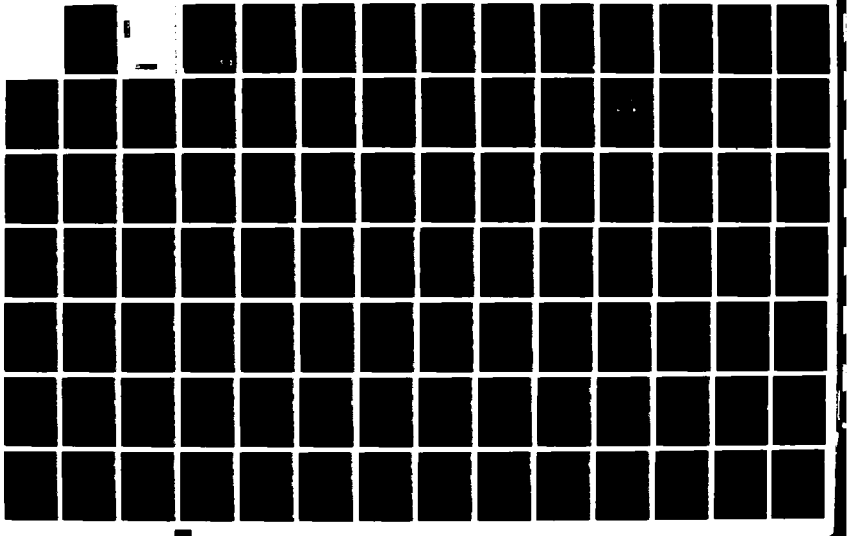
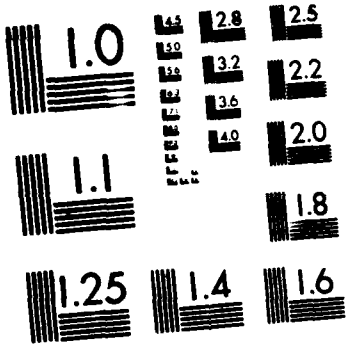
AD-A192 001 STATUS REPORT ON SPEECH RESEARCH(U) HASKINS LABS INC 1/2
NEW HAVEN CT M STUDDERT-KENNEDY SEP 87 SR-91(1987)
PHS-HD-01994

UNCLASSIFIED F/G 5/7 NL





2

Status Report on Speech Research

SR-91
July-September 1987

Haskins Laboratories
New Haven, Connecticut 06511

Distribution Statement

Editor-in-Chief
Michael Studdert-Kennedy

Editor/Book Designer
Nancy O'Brien

Text Processor
Yvonne Manning

This publication repo. the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor-in-Chief at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut 06511-6695

Acknowledgment

The research reported here was made possible in part
by support from the following sources:

National Institute of Child Health and Human Development:

Grant HD-01994
Contract NO1-HD-5-2910

National Institutes of Health:

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8520709

**National Institute of Neurological
and Communicative Disorders and Stroke**

Grant NS 13870
Grant NS 13617
Grant NS 18010
Grant NS 24655

Investigators

Arthur Abramson*
Peter J. Alfonso*
Thomas Baer
Fredericka Bell-Berti*
Catherine T. Best*
Geoffrey Bingham**
Gloria Borden*
Susan Brady*
Catherine P. Browman
Etienne Colomb†
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Janet Fodor*
Anne Fowler**
Carol A. Fowler*
Ram Frost‡
Louis Goldstein*
Vicki L. Hanson*
Katherine S. Harris*
Leonard Katz*
J. A. Scott Kelso*
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Hiroshi Muta†
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Douglas Whalen

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Yvonne Manning
Bruce Martin
Vance Maverick*
Nancy O'Brien
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Joy Armson
Dragana Barac
Eric Bateson
Suzanne Boyce
André Cooper
Margaret Hall Dunn
Elizabeth Goodell
Joseph Kalinowski
Bruce Kay
Rena Arens Krakow
Deborah Kuglitsch
Hwei-Bing Lin
Katrina Lukatela
Paul Macaruso
Harriet Magen
Diana Matson
Gerald W. McRoberts
Pamela Mermelstein
Sheri L. Mize
Maria Mody
Mark Pitt
Nian-qi Ren
Lawrence D. Rosenblum
Arlyne Russo
Richard C. Schmidt
Jeffrey Shaw
Caroline Smith
Robin Seider Story
Mark Tiede
Karen Zaltz

*Part-time
**NIH Research Fellow
†Fogarty International Fellow, Lausanne, Switzerland
‡Postdoctoral Fellow, Hebrew University, Israel
†Visiting from University of Tokyo, Japan

Contents:

Status Report on

Speech Research

SR-91 July-September 1987

Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants

Catherine T. Best,[†] Gerald W. McRoberts,[††] and Nomathemba M. Sithole,[†††]

The language environment modifies the speech perception abilities found in early development. In particular, adults have difficulty perceiving many nonnative contrasts that young infants discriminate. The underlying perceptual reorganization apparently occurs by 10-12 months. According to one view, it depends on experiential effects on psychoacoustic mechanisms. Alternatively, phonological development has been held responsible, with perception influenced by whether the nonnative sounds occur allophonically in the native language. We hypothesized that a phonemic process appears around 10-12 months, which assimilates speech sounds to native categories whenever possible; otherwise, they are perceived in auditory or phonetic (articulatory) terms. We tested this with English-speaking listeners, using Zulu click contrasts. Adults discriminated the click contrasts; performance on the most difficult (80% correct) was not diminished even when the most obvious acoustic difference was eliminated. Infants showed good discrimination of the acoustically-modified contrast even by 12-14 months. Together with earlier reports of developmental change in perception of nonnative contrasts, these findings support a phonological explanation of language-specific reorganization in speech perception.

INTRODUCTION

Infants in the first half-year or so of life discriminate most speech sound distinctions with which they have been tested, including sounds that are not contrasted phonologically in the language of their environment, i.e., are not used to specify differences in word meanings, such as the English /r/-/l/ contrast that is irrelevant in Japanese (Aslin, Pisoni, Hennessy, & Perey, 1981; Best, 1984; Jusczyk, 1984; Lasky, Syrdal-Lasky, & Klein, 1975; Trehub, 1976; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984a). In some cases, the language environment may instead facilitate infants' initially poor discrimination of certain other nonnative contrasts, such as English /s/-/z/ (e.g., Aslin & Pisoni, 1980; Eilers, Gavin, & Oller, 1982; Eilers, Gavin, & Wilson, 1979; Eilers & Minifie, 1975; Eilers, Wilson, & Moore,

1977; Streeter, 1976). Adults, on the other hand, typically discriminate all native contrasts, but have difficulty discriminating nonnative contrasts (e.g., Abramson & Lisker, 1970; Goto, 1971; MacKain, Best, & Strange, 1981; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975; Singh & Black, 1966; Tees & Werker, 1984; Trehub, 1976; Werker et al., 1981; Werker & Logan, 1985; Werker & Tees, 1984b). The present report is concerned in particular with understanding developmental change in perception of early-discriminated nonnative speech contrasts.

The language environment clearly influences developmental speech perception. But is the influence due simply to differential auditory exposure or does it derive instead from the linguistic experience of acquiring and using the phonological system of the native language? In line with the latter suggestion, Eimas (1978) suggested that the speech perception abilities of infancy become reorganized by adulthood as a function of specific linguistic experience. Werker and her colleagues demonstrated that the perceptual reorganization has already occurred by 4 years of age (Werker & Tees, 1983) and, in fact, appears to take place around 10-12 months of age (Werker et al., 1981; Werker & Tees, 1984a; but see also Burnham, 1986). They also proposed a linguistic account of the reorganization—that infants presumably shift developmentally from perception of speech contrasts in pre-phonological terms, that is, based on their acoustic properties (physical characteristics such as frequency components, silent gaps, noise bursts) and/or phonetic properties (characteristics of the way the sounds are articulated), to perception of them in terms of contrasts that occur in the phonological system of their language (see also Jusczyk, 1982; MacKain, 1982), referred to as phoneme contrasts. Such a developmental change would seem well-suited to the infant's first steps in receptive and productive acquisition of words near the end of the first year (e.g., Lenneberg, 1967; Stark, 1980).

An alternative proposal, however, appeals to an auditory or psychoacoustic influence of experience, by which exposure to and processing of the acoustic properties of native speech sounds causes some change in the responsiveness or "tuning" of the listener's auditory system. The psychoacoustic proposal states that discrimination of early-distinguished nonnative contrasts declines due to lack of auditory experience with sounds that do not appear in native phoneme contrasts (e.g., Aslin & Pisoni, 1980). We refer to this as the "auditory experience" argument. Note that in either the psychoacoustic or the linguistic hypothesis, the perceptual change need not be permanent, and may instead involve shifts in attentional mechanisms (e.g., Aslin & Pisoni, 1980; MacKain et al., 1981).

Any explanation of language-specific effects on speech perception should take into account the relationship of phonetic properties to phonemic contrasts, which are defined by particular combinations of phonetic features. To illustrate, /b/-/p/ (represented at the phonemic level) share the phonetic features of bilabial closure and stop or obstruent manner of production, but they differ in voicing. The phones (phonetic representations that don't specify their phonological status in a given language) in this contrast, in utterance-initial position in English as in <bat>-<pat>, are [b]-[p^h] or [p]-[p^h]. That is, the phoneme /b/ typically has the phonetic feature either of voicing that occurs slightly before or simultaneous with the bilabial release ([b]) or of a short delay in voicing ([p]), and the phoneme /p/ has the phonetic feature of aspiration ([p^h]) during the longer delay of voicing (Goldstein & Browman, 1986). MacKain (1982) and Tees and Werker (1984) point out that auditory experience with a given pair of phones is not ruled out simply because they fail to appear in a native phoneme contrast. Phone differences that are put to phonological use in some foreign language, but not in the native language, may nonetheless occur as allophones, or within-category phonetic variations, of some native category (although many phones that appear in nonnative phoneme contrasts never, or only

very rarely, occur as allophones of native phonemes). In those cases of allophonic representation, auditory experience would occur without corresponding to phonological relevance. For example, some Arabic languages such as Farsi include a phonological contrast between the voiced velar stop /g/, and a voiced uvular stop /G/ that does not contrast with /g/ in English (Maddieson, 1984). Yet certain cases of the phone [G] can occur in English as allophonic variants of the phoneme /g/ in the context of back vowels such as /u/ and /a/, which cause /g/ to be articulated farther back than in the context of front vowels such as /i/.

The linguistic proposal has thus been extended to address nonnative contrasts that occur as allophonic variants in the native language. Werker and colleagues (1981, 1984a) argue that developmental decline in discrimination occurs for contrasting sounds that do not partake in native phonological contrasts, even if they occur as native allophonic variants. We refer to this as the "specific phonological relevance" hypothesis. A problem for this hypothesis, as well as for the "auditory experience" hypothesis, is that there appears to be some variability in the degree of perceptual reorganization for various nonnative contrasts. Recent findings reveal relatively good discrimination of nonnative Hindi aspirated voiced vs. voiceless stops /d^h/-/t^h/ (Werker & Tees, 1984b) and Farsi /g/-/G/ by English-speaking adults (Polka, 1987), and of nonnative /w/-/r/ by Japanese-speaking adults (Best, MacKain, & Strange, 1982).

To address this problem, Tees and Werker (1984) proposed that allophonic variants may provide the listener with experience that maintains some discrimination of phonetically similar nonnative phones, an "allophonic experience" hypothesis. Indeed, English-speaking adults are able to discriminate, especially after perceptual training and/or reduced memory load, nonnative contrasts in which the members occur as allophonic variants in English, such as Hindi [d^h]-[t^h] and the utterance-initial prevoiced vs. voiceless unaspirated [b]-[p] found in Spanish and other languages (Tees & Werker, 1984; Werker & Logan, 1985; Werker & Tees, 1984b; see also Carney, Widin, & Viemeister, 1977; Pisoni, Aslin, Perey, & Hennessy, 1982; Pisoni & Lazarus, 1974). In contrast, listeners have persistent difficulty with many nonnative distinctions in which one or both members fail to appear allophonically in English (Tees & Werker, 1984). However, this explanation does not distinguish between nonnative contrasts whose members are allophonic variants of a single native phonemic category versus those that may be variants of different native categories. The Farsi /g/-/G/ contrast is an example of the former case for English-speaking listeners. As an example of the latter, in intervocalic position the apical flap [ɾ] and the alveolar trill [r] are contrasted phonologically in Spanish but not English; however, [ɾ] is an allophonic variant of American English /t/ and /d/, and [r] is a British variant (Scottish accent) of American English [ɹ]. Note, though, that [ɾ]-[r] are not contrasted phonologically either in American English, which does not use the trill as an allophone of /r/, nor in British English, which employs both [ɾ] and [r] as allophonic variants of /r/. The difference in allophonic status of /g/-/G/ vs. /ɾ/-/r/ should lead to better discrimination of the latter than the former contrast by English-speaking listeners, if variability for nonnative contrasts does indeed depend on experience with allophones. Yet just the opposite pattern has been found (Oller & Eilers, 1983; Polka, 1987). Moreover, anecdotal evidence from English listeners suggests that certain other nonnative contrasts are quite discriminable even though neither element occurs as an allophonic variant in English, such as the glottalized stop distinction /k'/-/t'/ found in Tigrinya and !Xoo (Maddieson, 1984). However, other very similar contrasts that also fail to appear allophonically, such as the glottalized velar-uvular stop distinction /k'/-/q'/ which is found in Thompson (a Northwest American

Indian language), are very difficult for English listeners to distinguish (Werker & Tees, 1984a, 1984b).

A psychoacoustic explanation of perceptual variability, that is, one that appeals to generalized rather than linguistically-specialized responses of the auditory system to various acoustic properties, might seem to handle such exceptions more easily. Burnham (1986) suggests that relatively discriminable nonnative contrasts are distinguished from poorly-discriminated ones because the former are psychoacoustically "robust" while the latter are psychoacoustically "fragile." Fragile contrasts are believed to be lost in infancy and remain difficult for adults even after perceptual training or reduction of memory demands. Robust contrasts, on the other hand, are presumably discriminated until at least 4-6 years, and are more amenable to perceptual training in adulthood. Although this psychoacoustic approach might account for discriminability of nonnative contrasts that do not occur allophonically, it also has shortcomings. Most important, to avoid tautology, the hypothesis would require the establishment of language-independent criteria, such as a description of the acoustic features that should be associated with either end of the robust/fragile dimension, and/or a hierarchy of the relative difficulty that some nonhuman species have with various speech contrasts.

In the present research, we hypothesized that for listeners who have acquired the phonological system of their native language (or have begun to do so), attention is focused, during speech perception, predominantly at the phonemic level. For simplicity's sake, we refer to this as *phonemic perception* (see also Werker & Logan, 1985). It entails the perceptual assimilation[1] of incoming speech sounds to the phonemic categories of the native language whenever possible. Assimilation may take place regardless of whether those sounds are native or nonnative, and regardless of whether they actually occur allophonically or are simply phonetically similar to some native category. Nonnative contrasts can be divided into four classes: those in which 1) the contrasting phones are assimilated as variants of a single native category ("single-category" assimilation); 2) the phones are assimilated as the opposing members of a native phonological contrast ("opposing-category" assimilation); 3) one member is better assimilated to a native category (more similar phonetically) than the other ("category-goodness difference" assimilation); and 4) both members are phonetically dissimilar from any native categories, and are therefore not assimilated ("non-assimilation").

In "single-category" assimilations, discrimination should be difficult for adults, even with perceptual training, and decline in discrimination should occur by 10-12 months of age. The last three classes should be discriminated by adults and older infants, but for different reasons. "Opposing-category" assimilations should be perceived as phonemic contrasts. "Category-goodness difference" assimilations should be perceived as a difference in "goodness of fit" for a native phoneme category. That is, although attention is primarily focused at the phonemic level, listeners should retain some sensitivity to within category phonetic articulatory variations that show differences in degree of match with the phonetic properties of the "ideal" category exemplar. Thus, the second and third classes, along with the first, involve perception of phonemically-relevant information. In contradistinction, "non-assimilated" contrasts should be perceived in terms of their auditory (acoustic or nonspeech properties) or phonetic (phonologically-neutral articulatory) characteristics.

Because of linguistic constraints on possible phonological oppositions, most of the contrasts of the world's languages naturally fall into the first three classes. Previous research, including Werker's studies of infants, has focused primarily on nonnative contrasts of the first class. Examples of "single-category" contrasts are Thompson /k'/-/q'/, which assimilate to English /k/ (Werker & Logan, 1985), and Spanish

intervocalic /r/-/ɾ/ (Oller & Eilers, 1983), which should assimilate to /t/ or /d/ for American English listeners, and to /r/ for British listeners. Perceptual difficulty with the latter contrast is problematic for the "allophonic experience" hypothesis. The second and third class are represented in studies that found relatively good discrimination, and effectiveness of perceptual training, in adults for some nonnative contrasts (e.g., Aslin et al., 1981; Best et al., 1982; Polka, 1987; Tees & Werker, 1984; Werker & Tees, 1984b). An example of an "opposing-category" contrast is /k'/-/t'/, which should be assimilable to English /k/-/t/. This presumably discriminable contrast would be troublesome not only for the "allophonic" and "auditory experience" hypotheses, since glottalized stops don't occur phonologically or allophonically in American English, but also for the robust-fragile psychoacoustic distinction, since the acoustic difference between the /k'/-/t'/ release bursts is likely analagous in magnitude to that found in the poorly discriminated Thompson /k'/-/q'/. "Category-goodness difference" assimilation is represented by the /g/-/G/ contrast that English speakers discriminate even though English does not contrast /g/-/G/ in any vowel context,[2] and the English /w/-/r/ contrast that Japanese discriminate even though [ɹ] does not occur in Japanese. The listeners in the former "category-goodness" example reported hearing a good English /g/ versus a foreign sounding one. In the latter, the Japanese subjects recognized /w/ as a good example of their native /w/, while /r/ was deviant from any native category. Neither example corresponds well to predictions of the "auditory experience" or the "specific phonological relevance" hypotheses.

The research reported here focused on "non-assimilable" nonnative contrasts. Specifically, we assessed whether English-speaking adults would show good discrimination for nonnative contrasts whose phonetic characteristics are highly dissimilar from any native categories, and that therefore are unlikely to be assimilated. We also tested whether discrimination of these presumably non-assimilable contrasts depends on auditory cues that might be assumed to have robust psychoacoustic effects, or rather on other auditory or phonetic (articulatory) differences that might be considered to be psychoacoustically fragile because of their similarity to certain perceptually difficult, early-lost nonnative contrasts. Finally, we tested whether infants show perceptual change at 10-12 months for such a contrast, as they do for "single-category" contrasts.

For this research, we used the click consonants of Zulu, which appear neither as phonemic contrasts nor as allophonic variants in English, nor do their phonetic-articulatory features correspond well to English phonemes. Although American listeners have typically experienced clicks produced as nonspeech "mouth sounds" or affectively-toned vocalizations,[3] this does not constitute linguistic experience.[4] The clicks in spoken Zulu are produced with a vowel context, and carry coarticulatory information as well as consonantal phonetic features such as voicing, nasalization, or glottalization. Nonspeech clicks have none of these phonetic characteristics. In Experiment 1, we predicted that American English-speaking adults would be well able to discriminate click syllable contrasts, since they should not assimilate the clicks to English phonemes.

EXPERIMENT 1

Zulu, a Bantu language, is one of a number of tone languages from southern Africa that employ click consonants, which are ingressive, unlike any English consonants. Clicks are produced by the formation of a suction chamber in the oral cavity followed by an abrupt release of the negative pressure (Catford & Ladefoged, 1968; Doke, 1926; Ladefoged, 1971, 1975) at the blade, tip, or side of the tongue, or at the lips (kissing sound) as in !Xoo, a Khoisan language (Ladefoged & Traill, 1984). Since the suction

involves velar occlusion of airflow, click release in Zulu also includes subsequent velar release at varying delays of voicing (Doke, 1926). Zulu has 15 clicks, distributed across three different places of articulation: apicodental ([ǀ]), palatoalveolar ([ǂ]), and lateral alveolar ([ǁ]). Each is produced with one of five additional phonetic features. There are two categories of nasalized clicks (voiced or voiceless unaspirated) and three nonnasalized voicing categories (voiced, voiceless unaspirated, and voiceless aspirated) (Catford & Ladefoged, 1968; Doke, 1926; Maddieson, 1984; Nyembezi, 1972). We could find no published reports on perceptual or acoustic studies of the Zulu clicks, although Doke (1926) has described articulatory properties of Zulu clicks, and Ladefoged and Traill (1984) have described the articulatory and acoustic properties of clicks in several Khoisan languages. Zulu has a moderate number of clicks: compare with !Xoo, which has 5 places of articulation and 16 possible phonetic accompaniments (e.g., voicing, nasalization, glottalization, velarization, or combinations thereof) that can be applied at each place (Ladefoged & Traill, 1984).

In the Zulu apical (apicodental) click, the tongue tip is released from the back of the upper front teeth. For the palatal (palatoalveolar) click, the tongue tip and blade are released in midline at the front of the hard palate, behind the alveolar ridge. The lateral (lateral alveolar) click is asymmetrical, with one side of the tongue released from the lateral portion of the alveolar ridge (see Doke, 1926; Ziervogel, Louw, & Taljaard,"1976). Thus, the place of articulation for the apical click is only roughly similar to that for /t/ in English, and actually more like the Hindi dental stop /t̪/. Nothing even roughly equivalent to the palatal or lateral places occurs in any English stop. The asymmetrical release of the lateral click is in fact a very uncommon feature in the world's languages. We restricted our tests to the voicing and place contrasts among the nonnasalized clicks, for which there are 18 minimal-pair contrasts of either place or voicing. Because we predicted good discrimination, we tried tc minimize procedural biases toward good performance. Therefore, we used an AXB discrimination procedure with relatively long interstimulus intervals (ISIs) of 1000 ms, rather than one with lower memory demands such as 2IAX or 4IAX (see Carney et al., 1977; Pisoni & Lazarus, 1974; Pisoni & Tash, 1974) and/or with short ISIs of 250 ms or less (Pisoni, 1973; Werker & Logan, 1985). The task was designed so that the matching items were different tokens of a click category rather than physically identical (see also Werker & Logan, 1985). Such a task should tap some degree of perceptual constancy for items within a phonetic category. The click syllables were matched across categories for general acoustic properties (e.g., pitch, loudness, duration) to minimize discrimination on the basis of phonemically irrelevant information. Moreover, the subjects were not given training on the clicks nor feedback on the practice trials, as had been done in other studies reporting above-chance nonnative speech discrimination (e.g., Pisoni et al., 1982; Tees & Werker, 1984; Werker & Tees, 1984b).

Method

Subjects

Nine college students were tested (7 female, 2 male; age range = 19-23 years). All were monolingual American English speakers with no previous exposure to Zulu or other click languages. None had any known hearing or language difficulties. Each was paid $8.00 for participation in a 1-1/2 hour test session.

Stimuli

The test stimuli were selected from naturally produced Zulu click + /a/ syllables recorded by TM, a native Zulu-speaking woman born and raised just south of the Mahlabathini section in the heart of Zululand, South Africa. The accent of people

from this region is considered by Zulus to contain the purest pronunciations of the clicks. TM read from a randomized list containing 20 repetitions of each of the 15 clicks. All syllables were produced with high tone. To insure the desired tonality, examples of bisyllabic imperative verbs with click + /a/ in word-initial position were given for each item on the sequence listing (see Table 1); only the first syllable of the words was spoken. TM was instructed to keep her productions as constant as possible throughout the sequence with respect to duration, loudness, and pitch contour. The utterances were recorded with a Sony T5D portable cassette tape deck, using a directional Audio Technica microphone.

TABLE 1. The Zulu Words Used for Recordings of Click + /a/ Syllables, and Their English Glosses.

CLICK SYLLABLE[a]	ZULU VERB[b,c]	ENGLISH GLOSS
/ǂa/	<<u>c</u>aca>	be clear
/ǂʰa/	<<u>ch</u>aya>	spread out (v.)
/gǂa/	<g<u>c</u>aba>	make an incision
/ŋǂa/	<n<u>c</u>ama>	give up
/ŋǂa/	<ng<u>c</u>ama>	feast (v.)
/ǁa/	<<u>q</u>ala>	start (v.)
/ǁʰa/	<<u>qh</u>aba>	snap fingers
/gǁa/	<g<u>q</u>aba>	paint (face) (v.)
/ŋǁa/	<n<u>q</u>aba>	refuse (v.)
/ŋǁa/	<ng<u>q</u>anga>	shake
/ʖa/	<<u>x</u>axa>	beat (v.)
/ʖʰa/	<-<u>xh</u>ala>	anxiety (n.)
/gʖa/	<g<u>x</u>atha>	stride (v.)
/ŋʖa/	<n<u>x</u>axa>	coax, urge
/ŋʖa/	<ng<u>x</u>ama>	be angry

[a]Represented in phonetic symbols (see Catford & Ladefoged, 1968; Ladefoged, 1975).

[b]The words are written in current Zulu orthography (which is based on the Roman alphabet), in which <c> corresponds to the apical place of articulation, <q>corresponds to the palatal place, and <x> to the lateral. An <h> following a click symbol indicates that it is voiceless aspirated, whereas letters preceding a click symbol indicate voicing (<g>), nasalized voicing (<ng>) or nasalized voicelessness (<n>). The voiceless unaspirated items are represented by the click place symbols alone. Thus, for example, the apical click syllables are written as follows: <ca> (voiceless unaspirated), <cha> (voiceless aspirated), <gca> (voiced), <nca> (nasalized voiceless), and <ngca> (nasalized voiced). Although the nasalized click syllables were recorded, they were not used in the present perceptual experiments.

[c]The speaker produced each of the underlined syllables with high tone, as it is normally spoken in word-initial position in these words (except <-xhala>, a suffix in which the syllable of interest is in initial position), which were provided as examples on the sequence list. The initial syllable in all items on the list (including -xhala) is produced with high tone. All items are imperative verbs, except <-xhala>. No bisyllabic imperative verbs beginning with <xha-> exist in Zulu; in fact, we were unable to find any bisyllabic words beginning with <xha-> using high tone.

The nonnasalized click syllables were digitized and stored on disk, using the PCM (Pulse Code Modulation) system of the VAX 11-780 computer at Haskins Laboratories. Author NMS (a native Zulu speaker) eliminated any tokens that were pronounced incorrectly or unclearly, or with an incorrect tone or vowel quality. From the remaining syllables, six exemplars of each category were selected for their

similarity in length, loudness, pitch, and vowel quality. Preliminary acoustic analyses verified that the selected syllables were physically similar (except for click properties--see second paragraph below). Although there was some degree of variation among the tokens in acoustic properties of the vocalic (vowel) portions of the syllables (e.g., F_0, contour, amplitude), it was found within as well as between categories, and there was much overlap in these vocalic acoustic properties between categories. The original duration of the selected syllables ranged from 232-310 ms, with a mean of 285 ms. The durations of the syllables were modified, by means of a software waveform editor on the Haskins VAX 11-780, so as to restrict the final range to 272-302 ms (M = 288 ms). This was accomplished by iterating or deleting individual pitch pulses from the steady portions of the vowels, and/or by adding or deleting small amounts of silence in the closure portion of voiceless items. In these cases, NMS verified that the editing had not distorted the phonetic properties of the syllables.

F_0 and formant frequency characteristics of the syllables were calculated by LPC analysis (ILS software), while VOT and durations of click bursts were measured by hand-marking and measuring the waveform in the waveform editor program. The results are shown in Table 2. The average F_0 contour across the vocalic portions of the items was nearly flat, the overall mean at vowel onset being 202 Hz, and at vowel offset being 197 Hz. However, F_0 contour varied slightly along the voicing dimension, due to differences in onset frequency between voicing categories (offset frequency did not differ noticeably). The voiceless aspirated items had a slightly higher starting frequency (M = 212 Hz) than the voiceless unaspirated (M = 204 Hz) or voiced items (M = 190 Hz); the former were slightly falling, whereas the latter two were slightly rising. The F_0 onset difference between the vocalic portions of the voiceless aspirated and voiced click syllables may be akin to that found between English voiceless and voiced stops (Haggard, Ambler, & Callow, 1970).

Acoustic properties of the clicks differ across places of articulation and voicing categories (see Table 2). Place categories differ in duration of click bursts only slightly. They are somewhat longer for lateral clicks (M = 52 ms) than for apical (M = 44 ms) or palatal clicks (M = 43 ms). Voiceless aspirated click bursts are slightly longer (M = 53 ms) than voiceless unaspirated clicks (M = 46 ms), the latter in turn being slightly longer than voiced clicks (M = 40 ms). Click amplitude at the peak of the burst varied systematically, being highest for the palatal (M= 52.7 dB signal/noise ratio) and lateral clicks (M = 51.4 dB), and lowest for the apical clicks (M = 39.7 dB). The spectral distributions of the clicks also differed, with the apical click bursts showing relatively greater energy in the high frequency range than the other two categories, and the palatals showing relatively greater energy in lower frequencies.

VOT was measured as the time between onset of the burst and onset of periodic voicing (Lisker & Abramson, 1964). VOT was longest for voiceless aspirated clicks, and shortest for voiced clicks. It should be noted that although we use the click voicing terminology recommended by phoneticians (Catford & Ladefoged, 1968; Doke, 1926), the VOT durations do not correspond well with the VOT measurements that have been reported for stop voicing categories that carry the same name. In fact, all three click voicing categories involve a lag between burst onset and voicing onset, which would be termed "voiceless" in phonetic descriptions of stop consonants. The lag, even for so-called voiced clicks, is due to the suction mechanism, which prevents release of the velar occlusion (and hence voicing onset) until after release of the click. The VOT durations of voiced clicks do not correspond well with stop voicing categories in English (Lisker & Abramson, 1964). Voiced click VOTs are longer (M =

TABLE 2. Acoustic Measurements of the Nonnasalized Zulu Click + /a/ Syllables.

	ACOUSTIC MEASURES:					
	F_0 onset[a]	F_0 nucleus[b]	F_0 offset	Click duration[c]	VOT[c]	Click amplitude[d]
CLICK CATEGORIES:						
Unaspirated Voiceless						
apical (/ǃa/)	203	205	199	43.3	62.9	40.55
	(185-217)[e]	(196-213)	(185-204)	(35-50)	(36-92)	(37.4-43.7)
lateral (/ǁa/)	204	200	198	51.7	70.8	50.73
	(196-213)	(189-204)	(189-208)	(40-60)	(40-91)	(46.5-54.2)
palatal (/ǂa/)	207	205	197	44.2	55.8	52.94
	(189-222)	(204-208)	(192-204)	(35-50)	(45-71)	(48.3-54.8)
Voiceless Aspirated						
apical (/ǃʰa/)	219	208	201	50	153.8	40.09
	(213-222)	(200-213)	(196-208)	(35-60)	(93-148)	(38.1-42.7)
lateral (/ǁʰa/)	210	190	195	60	143.2	53.5
	(182-233)	(192-208)	(185-204)	(55-65)	(134-150)	(49.6-56.3)
palatal (/ǂʰa/)	208	206	198	47.5	121.4	52.47
	(196-222)	(200-217)	(189-204)	(45-50)	(105-140)	(50.7-54.9)
Voiced						
apical (/gǃa/)	194	190	200	40	33.1	38.3
	(189-196)	(185-196)	(189-204)	(35-50)	(29-39)	(35.6-40.7)
lateral (/gǁa/)	189	187	196	43.3	34.7	49.96
	(182-196)	(182-189)	(185-200)	(40-50)	(37-43)	(46.5-55.0)
palatal (/gǂa/)	186	188	191	37.5	31.1	52.7
	(175-196)	(182-192)	(189-192)	(35-45)	(22-40)	(48.3-54.8)

[a]Shown in Hz. Measured at vocalic onset.

[b]Measured at vowel nucleus, approximately 80 ms from syllable offset.

[c]Shown in ms.

[d]Shown in dB gain (signal/noise ratio) for a 10 ms window at amplitude peak of click burst.

[e]Range of values is shown in parentheses.

+31 to +35 ms) than those associated with voiced stops (M = -102 to +21 ms) but shorter than those found in English voiceless stops (M = +58 to +80 ms). The voiceless unaspirated clicks have VOTs (M = +56 to +71 ms) corresponding to English voiceless stops (but the latter are usually aspirated in initial position). The voiceless aspirated clicks have VOT values (M = +121 to +154 ms) far longer than those associated with English voiceless stops. Thus, none of the click voicing categories corresponds well to the acoustic and phonetic properties associated with English stop consonant voicing contrasts.

Procedure

Subjects completed an AXB discrimination test including comparisons for each of the 18 minimal-pair contrasts (see Table 3). Testing was conducted in a sound-attenuated room, with stimuli presented at a comfortable listening level (approximately 75 dB SPL) over Sennheiser HD230 headsets to groups of 2 to 4 subjects. The stimuli were played out from an Otari MX5050 BQ-II tapedeck.

TABLE 3. The Minimal-pair Contrasts among the Nonnasalized Zulu Click Syllables.

	CONTRAST TYPE: PLACE OF ARTICULATION		
MINIMAL CONTRAST:	apical vs. palatal	apical vs. lateral	palatal vs. lateral
FEATURE CONTRAST:			
VOICING CATEGORIES			
voiceless unaspirated:	/ɳa/-/ʗa/	/ɳa/-/ʖa/	/ʗa/-/ʖa/
voiceless aspirated:	/ɳʰa/-/ʗʰa/	/ɳʰa/-/ʖʰa/	/ʗʰa/-/ʖʰa/
voiced:	/gɳa/-/gʗa/	/gɳa/-/gʖa/	/gʗa/-/gʖa/
	CONTRAST TYPE: VOICING		
MINIMAL CONTRAST:	voiceless, aspirated vs. unaspirate	voiced vs. voiceless aspirated	voiced vs. voiceless unaspirated
FEATURE CONTRAST:			
PLACE CATEGORIES			
apical:	/ɳa/-/ɳʰa/	/gɳa/-/ɳʰa/	/gɳa/-/ɳa/
palatal:	/ʗa/-/ʗʰa/	/gʗa/-/ʗʰa/	/gʗa/-/ʗa/
lateral:	/ʖa/-/ʖʰa/	/gʖa/-/ʖʰa/	/gʖa/-/ʖa/

The AXB discrimination test contained 36 blocks of 12 trials each, randomized within blocks. Three stimuli were presented on each trial, and the subjects indicated on a check-off sheet whether the middle stimulus (X) was from the same category as the first (A) or the third (B) stimulus. On the same sheet, they circled a number from 1-4 for each trial to indicate their confidence in their answer (where 1 = simply guessing and 4 = very sure). The ISIs within trials were 1000 ms. The intertrial intervals (ITIs) were 6 s, and the interblock intervals (IBIs) were 10 s. Each test block was restricted to a single contrast; there were two test blocks for each of the 18 contrasts, one in the first half of the test and one in the second half. Subjects were given a 10-minute break between the first and second half of the test.

On each trial, the middle item was a non-identical token from the same category as either the first item (A) or the third item (B). We refer to this procedure as name-identity AXB discrimination (see also Werker & Logan, 1985). Subjects were told that on each trial, the first and third items were always from different speech sound categories, even if they didn't sound so, and that the middle item was from the same category as the first or third. They were given 12 practice trials without feedback, which ranged across the 18 contrasts.

After the end of the test, the subjects completed posttest questionnaires, asking them to describe the properties of the syllables they had used to base their discriminations upon, and how easy they had found the task.

Results

The subjects' confidence ratings were relatively high: a rating of 4 ("very sure") was indicated for 37% of the trials, and a rating of 3 ("sure") for 36%. The average rating across the test was 3.04.

The data for percent correct performance were entered into a within-subjects analysis of variance (ANOVA), in which the second and third factors were nested within the first factor: the design was 2 (Contrast Type: Place Contrast, Voicing Contrast) x 3 (Feature Category: voiceless unaspirated, voiceless aspirated, and voiced for Place Contrasts; apical, lateral, and palatal for Voicing Contrasts) x 3 (Minimal Contrast: voiceless unaspirated vs. voiced, voiceless aspirated vs. voiced, and voiceless aspirated vs. unaspirated for Voicing Contrasts; apical vs. lateral, apical vs. palatal, and lateral vs. palatal for Place Contrasts). Table 4 illustrates the design and displays the percentage of correct discriminations on each of the 18 click contrasts. Table 5 lists the significant effects of the ANOVA. The main effect for Minimal Contrast (within Contrast Type) indicates that performance was higher on the voiced vs. voiceless aspirated distinction than on the other two voicing distinctions, and that performance was higher on the place of articulation contrasts that included the palatal click than on those that did not (apical vs. lateral). The Feature Category x Minimal Contrast (Contrast Type) interaction revealed that regardless of voicing category, performance was lower for the apical vs. lateral contrast than for the contrasts involving the palatal clicks, and that the voiced vs. voiceless aspirated distinction was easiest regardless of place of articulation. While performance on the other two voicing distinctions was somewhat lower, it did not differ at either the palatal or the lateral places. However, the voiceless aspirated vs. voiceless unaspirated distinction at the apical place was the most difficult voicing distinction overall. The main effects of Contrast Type and of Feature Category (Contrast Type) were nonsignificant. There was no generally greater ease with voicing contrasts than with place contrasts.

TABLE 4. Mean Percent Correct Performance on Discrimination of the Minimal-pair Click Contrasts.

	CONTRAST TYPE: PLACE OF ARTICULATION		
MINIMAL CONTRAST:	apical vs. palatal	apical vs. lateral	palatal vs. lateral
FEATURE CONTRAST VOICING CATEGORIES			
voiceless unaspirated:	97.7	80.6	95.8
voiceless aspirated:	97.2	82.9	94.4
voiced:	92.6	86.6	96.8
	CONTRAST TYPE: VOICING		
MINIMAL CONTRAST:	voiceless aspirated vs. unaspirated	voiced vs. voiceless aspirated	voiced vs. voiceless unaspirated
FEATURE CONTRAST: PLACE CATEGORIES			
apical:	82.4	99.1	88.0
palatal:	88.0	98.2	89.4
lateral:	89.4	97.7	89.4

TABLE 5. Significant ANOVA Effects, Experiment 1.

	df	F	p
FACTORS:			
Minimal Contrast (w/in Contrast Type):	4, 32	17.95	.0000
Feature Category X Minimal Contrast (w/in Contrast Type):	8, 64	2.40	.025

Discussion

The results indicate that neither a lack of experience hearing clicks in spoken English, nor their phonological irrelevance, had a negative effect on click discrimination. They are consistent with our prediction that discrimination should be easy for nonnative contrasts that cannot be assimilated to English phoneme categories.

Experience listening to clicks as nonspeech (see Footnotes 2 and 3) might also explain the maintenance of perceptual sensitivity, however. This possibility is weakened by observations of other nonnative contrasts for which highly similar nonspeech experience should be relevant. For example, the intervocalic Spanish trill vs. flap contrast /ɾ/-/r/([r]) is assimilated to English /t/ or /d/, or to /r/([ɹ]) by English-speaking adults, and is apparently lost by the second half-year of life in English-learning infants (Eilers et al., 1982; Oller & Eilers, 1983). Yet [r] is similar to the rolling tongue-trill that is used in infants' vocal play, in children's imitations of airplane or car sounds, and in mimicry of the cat's purr, while [ɾ] is a common allophonic variant of American English /t/ or /d/ (intervocalically) or of British /r/.

Another potential explanation might be that the clicks are psychoacoustically robust, and thus resistant to decline in discriminability (Burnham, 1986). Although the very fact that the clicks are easily discriminated fits Burnham's definition of psychoacoustic robustness, one would prefer to use independent criteria. For example, Burnham hypothesized a correlation between the robustness of a contrast and its representation in world languages. While the click contrasts are relatively rare, this is not a serious problem for the psychoacoustic argument. The linguistic distribution of a contrast would presumably also be influenced by other factors, such as articulatory ease and/or sociocultural forces. Although robustness may be necessary for widespread adoption of a contrast, it is not sufficient.

The click distinctions might nonetheless satisfy other criteria for robustness, which might explain the variations in discrimination among the click voicing distinctions and among the place of articulation distinctions. Burnham (1986) argues that at least some non-English stop voicing distinctions are psychoacoustically robust (e.g., prevoiced [b] vs. voiceless unaspirated [p]). Indeed, nonnative stop voicing contrasts are relatively amenable to training (e.g., Aslin & Pisoni, 1980; Werker & Tees, 1984b). The voiced vs. voiceless aspirated click contrast, which yielded the highest discrimination performance in the present results, would certainly seem to be psychoacoustically more robust than the others, given that it involves the largest VOT separation. Such a psychoacoustic explanation would be compatible with our hypothesis that non-assimilable contrasts should be discriminated on the basis of their auditory (or phonetic) properties.

Compared with voicing, nonnative place of articulation contrasts have generally proven more difficult perceptually, and resistant to training (e.g., Tees & Werker, 1984; Werker & Tees, 1984b). Yet the click place contrasts were also quite easy to discriminate. Allophonic considerations cannot account for the pattern of click place discriminations. The presence of a phonetic feature in the apical click that is roughly similar to one found in English /t/ did not provide special perceptual aid. Instead, discrimination performance was best for place distinctions involving the palatal click, which differs in place of articulation from any English phoneme. Thus, the acoustic properties of the stimuli played a larger role in place discriminations than did their phonetic properties. The amplitude variation across the three click places is the most obvious acoustic difference. Specifically, the palatal place is associated with the highest amplitude click burst. The apical place, although most phonetically similar to English, was associated with the lowest performance and the lowest amplitude click burst.

The psychoacoustic approach thus appears to handle discrimination performance on click distinctions. However, it should also be able to explain why performance on these contrasts reached higher levels than it has on other robust contrasts identified by Burnham. Although the clicks differ across places of articulation in their spectral distributions and amplitudes, and slightly in their durations, variations in burst amplitude and spectrum also differentiate the Hindi retroflex vs. dental stop (/ʈ/-/t̪/) contrast (Tees & Werker, 1984; Werker et al., 1981; Werker & Logan, 1985; Werker & Tees, 1984a, 1984b). Those authors found the retroflex-dental stop discrimination to be very difficult for English-speaking adults even after training, although low memory-demand conditions (unlike the conditions of the present study) did lead to improved performance. Similarly, although the bursts of the Thompson /k'/ vs. /q'/ contrast appear to differ somewhat in amplitude and duration, that place contrast is also difficult for English-speaking adults, even with instructions and/or training (e.g., Werker & Tees, 1984b). Listeners indicate hearing both members of the Hindi /ʈ/-/t̪/ contrast as /d/, while hearing both members of the Thompson contrast as /k/ (Werker & Logan, 1985). It is this phonemic influence for the Hindi and Thompson contrasts, we argue, that is responsible for their poor discriminability. We suggest that the high performance on the Zulu clicks occurred because no such phonemic influence appears to have operated for them, thus permitting subjects more direct perceptual access to their auditory (nonspeech) or phonetic (articulatory) properties. The implication that phoneme perception may supercede perception of purely auditory or phonetic information in the signal is not new. It is supported by research on categorical perception of speech sounds in general (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), as well as by perceptual constancy (e.g., Kuhl, 1980) and phonetic trading relations in perception of phoneme categories (e.g., Best, Morrongiello & Robson, 1981), and recent demonstrations that speech perception takes precedence over auditory perception of the same signal (e.g., Whalen & Liberman, 1987).

Subjects' answers on the posttest questionnaires indicated a virtual failure to assimilate the clicks to English phonemes. All subjects stated that they relied on auditory (nonspeech) properties of the sounds ("clicks," "plops," "pops," "percussion instruments," "water drip," "finger snap," "clap") when discriminating the syllables. Interestingly, several subjects also indicated relying on articulatory (phonetic) differences ("tongue popping," "tongue clucking," "sounds coming from different areas of the mouth"). Although some also thought there might be secondary vowel-quality, intonational, or loudness differences between the syllables, they indicated that these were small and difficult to differentiate. Only two subjects related any of the sounds to English consonants, and both indicated that they were only able to use these consonantal associations for a couple of test blocks. One referred to <d-t>

differences (most likely associated with the apical voiced vs. voiceless aspirated distinction) and both referred to <k-g> differences (most likely associated with the lateral voiced vs. voiceless aspirated distinction).

We believe that the very high levels of performance in Experiment 1 reflect a perceptual focus on the auditory and/or phonetic properties of the clicks. However, this may not occur solely because their acoustic differences are psychoacoustically robust, but rather because a failure to assimilate the clicks to English phonemes results in a perceptual focus on their nonphonemic properties. The amplitude differences among the click place contrasts would seem the most likely source of a robust psychoacoustic difference. Therefore, in Experiment 2, we tested whether adults would still discriminate a click place distinction after the click amplitudes were equated, on the basis of the remaining acoustic differences.

EXPERIMENT 2

In this experiment we compared American listeners with Zulu listeners. This allowed us both to determine whether amplitude modifications of the clicks had distorted crucial phonetic properties of the syllables according to native listeners, and whether differential linguistic experience influenced discrimination on the basis of the remaining acoustic properties. To provide the best chance of observing a developmental reorganization in click discrimination by infants (see Experiment 3), as would be predicted by Werker's findings (Werker et al., 1981; Werker & Tees, 1984a), we chose the place contrast on which adults had shown the lowest performance in Experiment 1, the voiceless unaspirated apical vs. lateral distinction. This click contrast is represented in phonetic symbols as [ǃ] vs. [ǁ], and the syllables are written in the Zulu orthography as <ca> vs. <xa>. In the original stimuli, the /ǁ/ click burst was higher amplitude than /ǃ/ on oscillographic tracings (see row a of Figure 1) and in the ILS analyses (Table 2).

Method

Subjects

Eight (4 male, 4 female) monolingual English-speaking college students (age range = 19-22 years) formed the English language group.[5] Six additional students (2 males, 4 females) formed the Zulu language group (age range = 19-36 years). All of the latter group had been born and raised in South Africa, but were currently enrolled in colleges in New England. Author NMS was one of the Zulu subjects. All in the Zulu group spoke English fluently. Three were from Zulu-speaking areas of South Africa, and had learned Zulu as their first language. The other three were from Xhosa-speaking areas and had learned Xhosa as their first language, but also spoke Zulu fluently. It should be noted that Zulu and Xhosa are very closely related, both being Bantu languages spoken by the Nguni peoples of South Africa. Speakers of one language can generally understand conversation in the other, although many vocabulary items are unique to one or the other language. The click system is identical to that of Zulu.

None of the subjects had any known hearing or language difficulties. Each received $4.00 for 30-45 min. of participation.

Stimuli

The amplitudes of the click bursts were equated across the two categories (in terms of dB gain levels at the peak) via software waveform editing, by reducing the amplitude of the /ǁ/ click bursts (but not the vocalic portion) and increasing that of the /ǃ/ clicks. The perceived loudness of the clicks was constant across the two

categories. According to author NMS, the modified /ɪ/ clicks sounded "wet" and the modified /ɜ/ clicks sounded somewhat attenuated or "swallowed," but the changes did not interfere seriously with their phonemic category membership.

The remaining acoustic properties of the syllables are listed in Table 6. The between-category distinction appears to be marked primarily by differences in F_1 and perhaps F_2 transitions (see Figure 1, rows b and c), and in spectral distribution of the clicks. The rising F_1 transitions are more rapid for /ɜ/ than /ɪ/, although the magnitude of the frequency excursion is similar. F_1 frequency at the beginning and asymptote of the transition is higher for /ɜ/. Both categories show a slightly rising F_2 transition, but the onset frequency is higher for /ɪ/. F_3 is nearly flat for both categories.

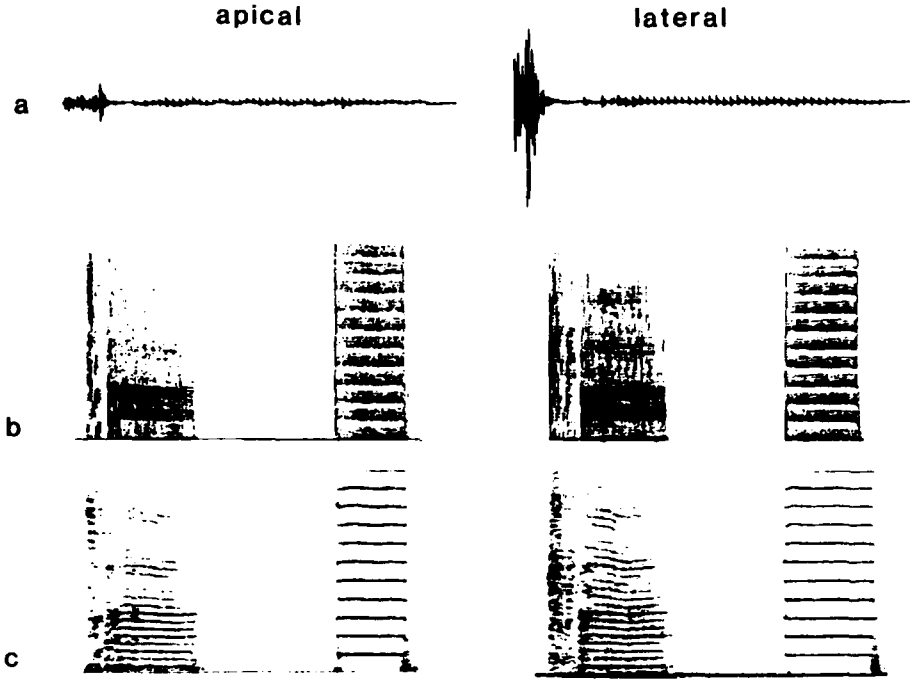


Figure 1. a) Oscillographic displays of an original (amplitude unmodified) voiceless unaspirated apical click (/ɪa/) and lateral click syllable (/ɜa/). b) Wide-band and c) narrow-band spectrograms of the amplitude-equated versions of those syllables (see Experiment 2).

Based on Discrete Fourier Transform (DFT) analyses of the click bursts, which present intensity x frequency information about brief portions of the signal, the tilt of the power spectrum differs at the highest-amplitude portions of the clicks (see Figure 2). The /ɪ/ (apical) clicks show a rising energy distribution with a secondary concentration of low frequency energy, whereas /ɜ/ clicks show both high and mid-range frequency peaks. There are also between-category differences in frequency characteristics of the onset and offset transients of the clicks. The onset transients are biased toward higher frequencies in /ɪ/. The offset transients for /ɪ/ show two distinct concentrations somewhat below the F_1 onset frequency and near the F_2 steady-state, while /ɜ/ shows offset transients with energy concentrations near F_3 and F_4.

The duration of the syllables and of the clicks alone differed only very slightly, and showed much overlap between categories. Closure durations were nearly identical, as were F_0 contour and level.

TABLE 6. Acoustic Measurements of the Amplitude-modified Voiceless Unaspirated Apical and Lateral Click Syllables.

	apical (/ǂa/)	lateral (/ǁa/)
ACOUSTIC MEASURES		
VOCALIC PORTIONS:		
F_1 onset	699 Hz (627-824)	795 Hz (507-995)
F_1 transition asymptote	1147 Hz (1113-1175)	1232 Hz (1037-1380)
F_1 transition change in frequency	448 Hz (329-529)	437 Hz (282-680)
F_2 onset	1557 Hz (1420-1666)	1427 Hz (1022-1571)
F_3 steady-state	2742 Hz (2601-2796)	2772 Hz (2642-2816)
F_0 onset	203 Hz (185-217)[a]	204 Hz (196-213)
F_0 at vowel nucleus[b]	205 Hz (196-213)	200 Hz (189-204)
F_0 offset	199 Hz (192-204)	198 Hz (189-208)
CLICKS:		
Frequency peaks	4655 Hz (4586-4736) 119 Hz (109-126)	4355 Hz (3971-4661) 2453 Hz (2138-2843)
Onset transient peaks	4557 Hz (4236-4810) 3248 Hz (1443-4023)	4529 Hz (4066-4661) 2485 Hz (2115-2739)
Offset transient peaks	1538 Hz (1322-1609) 440 Hz (431-454)	4476 Hz (4275-4741) 2420 Hz (2190-2509)
DURATIONAL MEASURES:		
Syllable length	286 ms (274-296)	293 ms (286-302)
Click duration	43 ms (36-54)	52 ms (39-64)
VOT	63 ms (36-91)	71 ms (40-91)
F_1 transition duration	64 ms (50-80)	30 ms (20-50)

[a]Range indicated by numbers in parentheses.
[b]Measured at approximately 80 ms from vocalic offset.

Procedure

The English language group was tested under the same experimental set-up as described for Experiment 1. The members of the Zulu group were tested, three at a time, in a quiet room near their college. The tests were presented to the latter group over the built-in speaker of a portable Sony T5D cassette tape deck at a comfortable listening level (approximately 75 dB SPL). All listeners completed the test(s) in a single session.

Both language groups completed a name-identity AXB discrimination test, of the same format as described for Experiment 1, except that it contained only trials with /ǁa/-/ǂa/ tokens and consisted of 6 blocks of 12 trials. For the English language group only, the first block of 12 trials served as a no-feedback practice set; their answers on this block were not scored. The Zulu language group first completed a standard identification test on the modified stimuli. This consisted of 5 blocked randomizations of the 12 stimuli, presented singly, with 2.5 s ISIs and 4 s IBIs.

Results Identification Task, Zulu Language Group

Overall, the Zulu listeners found the modified syllables to be acceptable tokens of apical and lateral voiceless unaspirated clicks. The group labeled the tokens correctly on 92% of the trials. Performance of author NMS was not noticeably different from that of the other Zulu listeners. Performance was somewhat higher for

/ʖ/ tokens (*M* = 98% correct) than for /ǀ/ tokens (86%), suggesting that the latter may have been somewhat less acceptable phonetically. However, this difference did not reach standard levels of significance according to ANOVA (p=.14). The Xhosa speakers had more difficulty with the /ǁ/ items (*M* = 79% correct), but only very slightly more difficulty with the /ʖ/ items (*M* = 96%), than did the Zulu speakers (*M* = 92% and 100%, respectively). Post-hoc t-tests of differences between the language subgroups on both comparisons failed to reach significance (p's > .20). However, the small n's restrict confidence in the null results of these statistical tests.

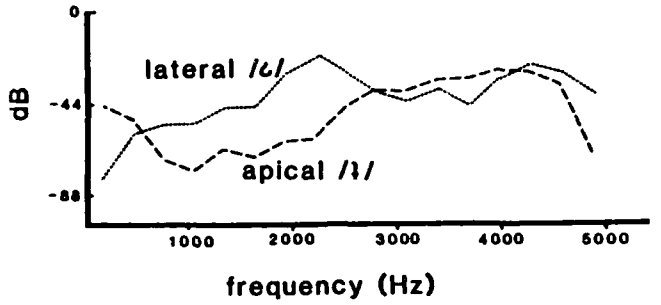


Figure 2. Spectral sections of the highest-amplitude portions of the click bursts from an amplitude-modified voiceless unaspirated apical (solid line) and lateral (dotted line) click syllable (see Experiment 2).

AXB Discrimination, Both Language Groups

The data for percent of correct answers on the AXB name-identity discrimination tasks were entered into a one-way ANOVA with Language Group as the between-subjects factor. For this purpose, the Zulu and Xhosa listeners were collapsed into a single Zulu language group. Although mean performance was somewhat higher for the Zulu language group (87% correct: 84% for Xhosas and 89% for Zulus) than for the English language group (78% correct), the difference was not statistically significant (p = .12). However, this null result is qualified, again, by the small n of the Zulu group, as well as by language group differences in listening conditions and in initial practice with the tokens.

Discussion

The results of Experiment 2 indicate that Zulu speakers can identify, and both Zulu and English speakers can easily discriminate, the voiceless unaspirated apical vs. lateral click contrast on the basis of acoustic distinctions other than click amplitude differences. Moreover, the English language group in the present study did virtually as well with these amplitude-modified stimuli (78% correct) as the listeners in Experiment 1 had done with the unmodified version of this contrast (80%), suggesting that amplitude variations did not play a large role in perception of the contrast in the earlier study. Further research involving parametric variations of the remaining acoustic cues (e.g., via digital resynthesis) in amplitude-equated click syllables could determine their relative importance in click perception by native and nonnative listeners.

The possibility for developmental reorganization in perception of Zulu clicks by English listeners, however, still remained. Although adults are well able to discriminate click distinctions, infants might show a decrement in discrimination of the amplitude-modified clicks at the same age as they have for other nonnative contrasts (e.g., Werker et al., 1981; Werker & Tees, 1984a). If so, it would indicate that

different processes underlie the responses of adults versus infants over 10-12 months of age to nonnative speech contrasts. For example, whereas adults' perception shows clear phonemic influences (i.e., assimilating the nonnative sounds to native phonemes where possible; otherwise focusing on auditory or articulatory information), infants' perception at 10-12 months may show simpler language-identity influences (i.e., maintaining attention only to sounds that are familiar in the language environment). In this scenario, infants would simply "tune out" or stop attending to differences between the unfamiliar, strikingly non-English, clicks by around 10-12 months. On the other hand, infants might perceive the clicks similarly to the adults, and thus continue to discriminate click contrasts even at 10-12 months and beyond.

EXPERIMENT 3

To test for these possibilities, we conducted a cross-sectional study of /ɬa/-/ʃa/ discrimination by 6-8 month, 8-10 month, and 10-12 month old infants for comparison with Werker's findings on other nonnative contrasts. To increase our opportunity to observe some perceptual decline for the click contrast, we added a fourth group of infants at age 12-14 months. We also tested discrimination for the native English stop place contrast, /ba/-/da/, as a control comparison. For this study, we adapted an operant-conditioning visual habituation procedure (Miller, 1983) (see Procedure). It should be noted that this procedure differs from the conditioned head-turn paradigm used by Werker and colleagues (Werker et al., 1981; Werker & Tees, 1984a). We chose to use the visual-fixation technique because it appears amenable to testing infants across a wider age range (at least from 2-13 months) than either the nonnutritive sucking technique (usable only between 1-4 months: e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) or the visually-reinforced head-turn technique (trainable only after 5-6 months: e.g., Eilers et al., 1979; Werker et al., 1981). We hoped it would be useful for future developmental studies.

Method

Subjects

The subjects were 40 infants from middle- and upper middle-class homes. A total of 62 infants was tested. Twenty-two infants were eliminated from the final data set for the following reasons: crying (8), inattention to the visual fixation target (6), equipment failure (4), experimenter error (1), inability of the observer to determine accurately the infants' fixations of the visual target (2), and performance levels that were more than 2 standard deviations beyond the means for the infant's age group (1). This left 10 infants each in the following age groups: 6-8 months (27-35 weeks), 8-10 months (36-44 weeks), 10-12 months (45-52 weeks), and 12-14 months (53-61 weeks). The mean age for each group was, respectively, 29.3 weeks, 39.2 weeks, 48.3 weeks, and 59.7 weeks. There were approximately equal numbers of boys and girls at each age. All were being raised in monolingual English environments. All had normal deliveries following normal full-term pregnancies. All infants were in good health at the time of testing; none was on medication. None had any known hearing problems.

For comparison purposes, the eight American adult subjects from Experiment 2 were also tested on the English and Zulu contrasts in the visual habituation procedure (see Procedure).

Stimuli

The Zulu discrimination test of the present experiment utilized the six amplitude-modified tokens each of /ɪa/ and /ɜa/ used in Experiment 2. The English discrimination test used six tokens each of /ba/ and /da/ produced by an adult male speaker of American English. The latter stimulus set was developed by Werker and colleagues (e.g., Werker et al., 1981) as their native-language control contrast, for which infants show good discrimination at all ages tested, as expected. The multiple tokens for each category thus provided a test of some degree of within-category perceptual constancy.

The stimuli for each test were recorded on a separate tape, such that a randomized sequence of the six items in one category appeared on one channel of the tape while the items in the opposing category appeared on the other channel, with exactly synchronous onsets. This permitted smooth switching between stimulus channels at the stimulus shift point during the test (see Procedure). The interstimulus intervals were 750 ms. Each tape contained approximately 45 minutes of continuously recorded stimulus presentations.

Procedure and Apparatus

All subjects were tested on both the Zulu contrast and the English contrast, during a single session. The infant-controlled visual habituation technique for assessing auditory discrimination was developed by Horowitz (1975) and adapted by Miller (1983) to test for perceptual constancy within categories. Generally speaking, in this paradigm an initial series of auditory stimuli (the "familiarization" set) is presented contingent on the subject's fixations of a projected visual pattern. Habituation to the auditory familiarization stimuli is indexed by a decrement in visual fixation to a criterion level, at which point a shift to a new set of auditory stimuli (the "test" set) occurs. A significant postshift recovery of fixation time to the same visual target indicates discrimination between the auditory familiarization and test stimuli.

Subjects were tested in a sound-attenuated room adjacent (and connected by one-way windows) to the control room from which the observer and experimenter conducted the session. The visual-fixation slide was back-projected (15 cm x 15 cm) through a translucent rectangular sheet of acetate slightly larger than the projected image, which was affixed in the center of a one-way window. The remainder of the window was covered with opaque black material except for a small peephole (invisible to the subject) by the lower right corner of the projected image. Two slides were used, one for each of the language tests. Each showed a 4 x 4 checkerboard, one of blue and white and the other of yellow and green (equated for brightness), in the center of which was the broken outline of a circle in orange or red, respectively.

A small booth in the testing room was attached to the top and sides of the projection window. The booth was covered inside with black felt, including the ceiling. The walls of the booth measured approximately 1-3/4 m high x 1 m wide; the opening was approximately 1 m wide. Each adult or infant subject was positioned in the booth, with eyes approximately 45 cm from the projected slide. Adult subjects were seated on a chair; infants sat either in an infant seat stabilized on a small table (younger infants) or in a highchair (older infants) inside the booth. Parents of the infant subjects sat quietly behind or to the side of the booth out of the infant's view, except when the infant would not tolerate the seat or chair. In the latter cases, the parent held the infant in a sitting or standing position within the booth, such that the infant could not see the parent's face. Parents were cautioned not to talk to or in any way distract their infants or bias their responses during the test sessions. During testing, the parents wore Sennheiser HD230 closed-model headsets through which music was played, to prevent them from hearing the audio stimuli.

An observer, who also listened to music over headsets, viewed the subject's fixations of the projected checkerboard (as judged by corneal reflection) through the peephole. The observer depressed a "looking" key whenever the subject fixated the checkerboard; there were also buttons to depress whenever an infant subject was crying or sleeping. This information was all recorded by an Atari 800 computer, which was programmed to end each visual fixation trial by closing a Gerbrands shutter on the slide projector whenever the subject looked away from the slide for more than 2 s. The Atari re-opened the shutter after a 1 s ITI, and a new trial began. The computer terminal displayed commands to an experimenter, who could not see the subject. These were commands to play the audio stimuli whenever the subject fixated the checkerboard, to stop audio presentations whenever the subject looked away from the slide, and to switch from the "familiarization" stimulus channel to the "test" channel when the computed habituation criterion was reached. The experimenter stopped the tape, and switched channels, only during the silent ISIs between syllables. The habituation criterion was based on Miller's (1983) formula: a fixation-time decrement of 50% or more on two consecutive trials, relative to the average of the two longest-duration trials for the familiarization phase of that test. The program calculated and updated the habituation criterion on every trial. After habituation had been reached and the stimulus shift had been signaled, the session then continued with "test" stimulus presentations, until the habituation criterion was met again. The computer automatically terminated the session at this point, or whenever any infant accumulated 30 s of crying or sleeping during a session.

The auditory stimuli were played to the subjects from the Otari tape deck, through a Kenwood amplifier and a specially constructed listening station that permitted easy switching of audio channels, and into a Jamo compact loudspeaker centered over the projection panel, above the ceiling of the booth. The stimuli were played at approximately 75 dB SPL.

Upon arrival at the laboratory, infant subjects were given a 5-10 minute period of acclimation while the procedures were briefly explained to the parent. Parents completed a permission form and a background questionnaire on family language background and on medical characteristics of the pregnancy and delivery. Each infant then participated in the two discrimination tests, one for Zulu (/ɬa/-/ʃa/) and one for English (/ba/-/da/). The order of test presentations was counterbalanced across infants, with approximately equal numbers of infants at each age in each test order. The infants were given a 5-10 minute break (or longer, if needed) between tests in order to minimize any carryover of habituation from the first to the second test.

The adults were given brief verbal instructions prior to the task, which indicated that there were both between- and within-category variations among the stimulus sets, and that they should listen as long as they wished (dependent on fixating the slide) until they felt familiar with the range of variation. Because the task is rather unusual for adults, all these subjects were tested with English first to familiarize them with the procedure.

Results

Infants

Three preliminary ANOVAs were conducted on the familiarization-phase data, to assess whether there were any important language-related differences in habituation. Significant effects of all ANOVAs are listed in Table 7. A 4 (Age) x 2 (Language) x 2 (Test Order) ANOVA was conducted on the data for total number of trials taken to reach the habituation criterion. The effects were all nonsignificant, indicating no systematic language-related variation on this measure of habituation. Another 4 x 2 x 2 ANOVA was conducted with the data on the cumulative fixation time

before reaching the habituation criterion. A significant Language x Test Order interaction indicated a larger cumulative fixation time during the familiarization phase of the first test for English (M = 61.03 s) than for Zulu (M = 46.15 s). Cumulative fixation during familiarization on the second test was lower and did not differ substantially between languages (Ms = 41.64 and 39.19 s, respectively). This pattern indicates that the infants preferred listening to the English syllables more than the Zulu during the first test, but that this difference disappeared by the second test, probably due to a general response decrement (though small) across the test session. A third 4 x 2 x 2 x 2 (Habituation Trials) ANOVA tested for differences in extent of habituation. The Habituation Trials effect indicated a significant difference between mean fixation for the two familiarization trials with highest fixation durations, versus the mean for the two trials just prior to stimulus shift, that is, significant habituation (see Figure 3).

TABLE 7. Significant ANOVA Effects for Experiment 3

	df	F	p
ANOVA EFFECTS, Infants:			
CUMULATIVE FIXATION DURING FAMILIARIZATION			
Language X Test Order	1, 32	4.50	.04
HABITUATION DURING FAMILIARIZATION			
Habituation	1, 32	53.41	.0000
PRESHIFT VS. POSTSHIFT MEANS			
Recovery (pre vs. post)	1, 32	37.00	.0000
Language	1, 32	3.07	.09
Language X Test Order	1, 32	16.73	.0003
Recovery X Age X Test Order	3, 32	3.24	.03
Recovery, Zulu alone	1, 32	33.55	.0000
Recovery, English alone	1, 32	10.21	.003
DIFFERENCE SCORES			
Age X Test Order	3, 32	3.24	.03
Age X Test Order, Zulu alone	3, 32	3.30	.03
ANOVA EFFECTS, Adults:			
PRESHIFT VS. POSTSHIFT MEANS			
Recovery (pre vs. post)	1, 6	24.25	.003

For the analyses on discrimination performance, the mean fixation time for the last two preshift trials, and the mean for the first two postshift trials[6] were computed, following Miller's (1983) approach for analyzing data collected in the visual-fixation techniques. A preshift-postshift comparison measures the occurrence and degree of stimulus discrimination as the extent of immediate postshift recovery in conditioned fixation. It is similar to the data analysis approach for the high-amplitude-sucking habituation technique (e.g., Eimas et al., 1971; Streeter, 1976). These data were first entered into a 4 (Age) x 2 (Sex) x 2 (Language) x 2 (Test Order) x 2 (Recovery: preshift vs. postshift) ANOVA (see means, Table 8 and Figure 3), to determine whether we could collapse across sex. The results indicated no systematic effects of sex, so the ANOVA was recomputed without Sex as a factor (see Table 7). The main effect of Recovery indicated that the postshift fixation times were

longer than the preshift times, that is, the change was discriminated. There were no other significant main effects in the overall ANOVA, although the Language effect approached significance, suggesting a trend toward higher fixation times to English (M = 4.94 s) than to Zulu (M = 3.78 s). However, note that this refers to the mean of preshift and postshift fixations. Therefore, it does not indicate language differences in discrimination, but rather may suggest something like a higher general interest level in English, as found for cumulative fixation during familiarization. Indeed, recovery was significant for each language separately (Table 7). There was no significant Language x Recovery interaction, indicating that discrimination was not consistently stronger for English than for Zulu.

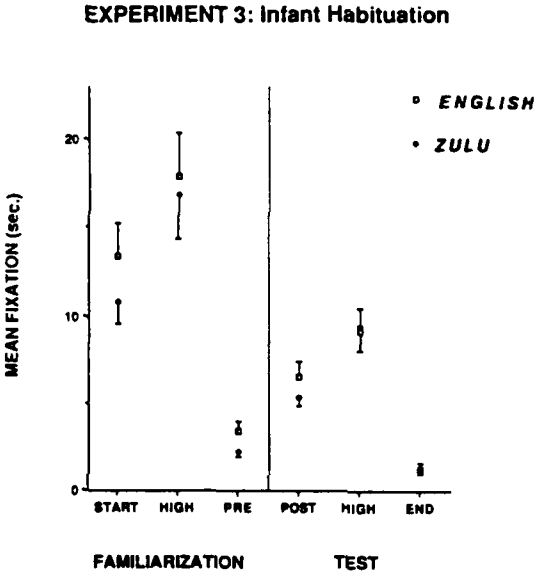


Figure 3. Habituation of infants' visual fixations during the familiarization phase, and dishabituation following the stimulus shift (test phase), for tests with the Zulu click and English stop contrasts. Each data point represents the mean fixation time, across ages, of a 2-trial block (START=first 2 trials; HIGH=2 trials with highest fixation times, for FAMILIARIZATION or TEST phase; PRE=the 2 trials immediately preceding the stimulus shift; POST=the first 2 stimulus-shift trials; END=last 2 trials of the test phase). The vertical bars extending from each data point represent standard error scores. It should be noted that standard habituation curves averaged across all trials of all infants could not be computed since the infant-controlled visual fixation procedure yields variable numbers of trials for individual subjects. Also note that due to rapid habituation, (as few as 4 familiarization or 4 test trials) in some infants, there is partial or total overlap in the test phase data for POST and HIGH, or in the familiarization phase data for START and HIGH, for approximately one-third of the subjects, on the Zulu and/or the English test.

There were two other significant interactions in the overall ANOVA. The Language x Test Order interaction showed the same pattern in the averaged preshift and postshift fixation times as had been found in the analysis of cumulative fixation during familiarization: higher values for English when it was tested first (M = 6.35 s) than for Zulu (M = 4.38 s), and lowest values for both when tested second (M = 3.14 and 3.25 s, respectively). This pattern did not reflect a difference in discrimination performance, though, since the Language x Test Order x Recovery interaction was not significant. It may instead reflect (again) an overall attentional preference for English sounds.

TABLE 8. Mean Fixation Times for Preshift and Postshift Trial Blocks, for Each Age Group on Each Language Test.

LANGUAGE:	English		Zulu	
RECOVERY:	preshift[a]	postshift[b]	preshift	postshift
AGE:				
6 months	5.09 (1.56)[c]	6.45 (1.61)	2.82 (0.57)	6.39 (1.07)
8 months	3.09 (0.87)	6.55 (2.13)	1.74 (6.47)	4.99 (1.23)
10 months	3.20 (0.94)	5.01 (1.04)	2.04 (0.60)	4.11 (0.84)
12 months	2.24 (0.61)	7.89 (2.50)	2.21 (0.67)	5.91 (0.84)
adults	6.50 (1.51)	20.13 (4.15)	9.95 (1.66)	15.45 (2.29)

[a]trial mean for last 2 trials prior to stimulus shift.

[b]means for first 2 trials of test stimulus presentations.

[c]standard error of the mean.

The Recovery x Age x Test Order interaction appeared to suggest that the postshift recovery in the oldest group (12-14 months) was greater for both languages when English rather than Zulu was tested first, but that the opposite held true for the three younger groups. In order to simplify the description of this three-way interaction, and verify the interpretation, another 4 (Ages) x 2 (Languages) x 2 (Test Orders) ANOVA was run on a measure of postshift recovery magnitude, calculated as difference scores (postshift fixation - preshift fixation). The Age x Test Order interaction was significant, since the difference scores are simple transformations of the data in the overall ANOVA. According to this simplified measure, also, the oldest group appears to discriminate better across both languages when tested on English first, whereas the younger three groups showed better discrimination when tested on Zulu first (see Figure 4). However, this interaction is essentially uninterpretable because Neuman-Keuls tests failed to reveal any significant pairwise differences among the data points, and also because the language difference was confounded with speaker gender differences. This Age x Test Order interaction was significant only for the Zulu tests.

Adults

To assess the discrimination performance of the adults on the visual habituation task, the mean of the last two familiarization trials (preshift) and the mean of the first two postshift trials were computed, as for the infants. These data were entered into a 2 (Sex) x 2 (Languages) x 2 (Recovery: preshift vs. postshift) ANOVA. The only significant effect was Recovery (Table 7), which indicated discrimination of the stimulus change across both languages. The Language x Recovery interaction was only marginal, suggesting a trend toward a greater discrimination of the English than of the Zulu "test" stimuli. Since the adults all received the same test order (English first), Test Order effects could not be assessed.

Zulu Click Discrimination

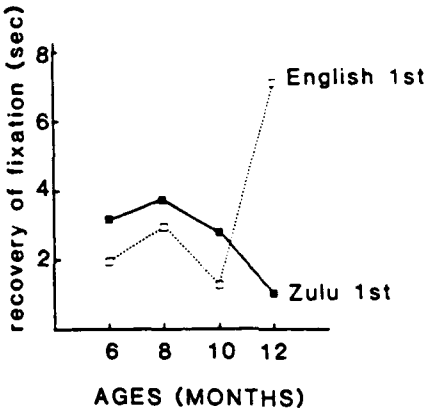


Figure 4. The Age x Test Order interaction found in the difference scores for the infants in Experiment 3. The ordinate represents the magnitude of fixation-time recovery in response to the stimulus change, computed as the mean fixation time on the first two postshift trials (test phase) minus the mean fixation time during the last two preshift trials (habituation criterion of the familiarization phase).

Discussion

The results indicate that in the visual habituation paradigm, both English-environment infants and adults discriminate the Zulu click category distinction between the amplitude-modified tokens of /ǀa/ and /ɓa/, as well as they do the English /ba/-/da/ contrast.

There was no significant infant age change in discrimination of the English versus the Zulu click contrasts, as would be predicted by both the "auditory experience" proposal (e.g., Aslin & Pisoni, 1980), and the "specific phonological relevance" proposal (Werker et al., 1981; Werker & Tees, 1984a). It should be recalled, in this context, that we employed a different procedure than did Werker and colleagues, as well as a different nonnative contrast. Moreover, whereas Werker's headturn procedure uses 1500 ms ISIs, we used 750 ms ISIs, possibly involving lower memory demands that could lead to improved performance. The possibility that differences between our infant findings and those of Werker and colleagues were caused by methodological differences needs to be tested. However, we suspect that the ISI differences are not crucial, since 750 ms is in the long ISI range for adults, which typically leads to phonemic-level perception of speech (Werker & Logan, 1985), and in fact other infant researchers have found significant language differences in infant speech perception using even shorter ISIs (400-600 ms) in the headturn procedure (e.g., Burnham, 1986; Eilers et al., 1977, 1979) and the sucking-rate habituation procedure (e.g., Streeter, 1976). Furthermore, the visual-fixation technique has been shown to be sensitive to age and stimulus differences in infants' perception of speech qualities (Miller, 1983).

The current results are consistent with the argument that the nature of the developmental change found by Werker and colleagues at 10-12 months for discrimination of (assimilable) nonnative contrasts is a transition toward perceiving speech sounds in relation to native language-specific phonemic contrasts. The similarity in the pattern of the infants' and adults' responses in the current visual habituation paradigm further suggests a commonality in their approaches to

the Zulu click contrasts. We suggest that for both age groups, this entails a failure to assimilate clicks to native phoneme categories.

GENERAL DISCUSSION

The overall pattern of results from the three experiments is consistent with our prediction that discrimination ability should remain high throughout infancy and in adulthood for nonnative contrasts that are unlikely to be assimilated to any native phonemic categories. This prediction was based on the reasoning that phonemic perception entails assimilation of nonnative speech sounds to native categories whenever possible, but that when they are not assimilated, perception focuses either on purely auditory or phonetic (articulatory) properties.

We argued earlier that neither the "psychoacoustic hypothesis" (e.g., Burnham, 1986) nor the "allophonic hypothesis" (e.g., Tees & Werker, 1984; Werker et al., 1981) alone could fully account for variation in developmental reorganization for the discrimination of nonnative contrasts. However, each may account for a different portion of the variation. Specifically, Tees and Werker suggested that allophonic experience maintains some degree of perceptual sensitivity for phonetically similar contrasts. Although that argument cannot account for the present findings with Zulu clicks, it may nonetheless apply to variations in performance on "single-category" and "opposing-category" contrasts, and possibly "category-goodness difference" contrasts (see *Introduction*): respectively, those that are assimilated to a single native phoneme, those that are assimilated to a native contrast, and those for which one member is better assimilated to a native category than is the other. On the other hand, Burnham's psychoacoustic proposal may apply most clearly to variations in "non-assimilable" contrasts, like the clicks. That is, psychoacoustic influences may be most apparent when perception is nonphonemic. They may also play a role in perception of the difference between the well-assimilated and the poorly-assimilated members of "category-goodness difference" contrasts.[7]

The concept of perceptual assimilation introduced in this paper, like the psychoacoustic robust/fragile distinction, calls for objective defining criteria.[8] We would offer that the likelihood and direction of assimilation (i.e., to which specific native phoneme[s]) of nonnative sounds should be predictable on the basis of the degree of similarity in phonetic-articulatory features between the nonnative item and the native categories. For example, although Thompson /k'/ and /q'/ are produced with non-English ejective manner, they share the feature of stop manner of articulation with English /k/, and the places of articulation for both Thompson phones occur in allophonic variants of English /k/. In contrast, Zulu /ɓ/-/ʄ/ share neither manner nor place with any English phoneme. Phonetic similarity criteria should derive from phonetic-articulatory features as established by phoneticians (e.g., Ladefoged, 1975).

Future research could assess more directly whether English-speaking adults discriminate the clicks on the basis of auditory or phonetic information. For example, Zulu but not English listeners would be expected to show such presumably speech-specific influences as trading relations between phonetic cues (e.g., Best et al., 1981) and vowel context effects (e.g., Mann & Repp, 1980). Werker and Logan's (1985) technique of determining different discrimination patterns for auditory, phonetic, and phonemic levels of speech perception could also be applied to cross-language group comparisons, and to comparisons between "non-assimilable" vs. "single-category" contrasts.

The current infant findings point out an important limitation of earlier findings of perceptual decline in discrimination of early-discriminated nonnative contrasts at around 10-12 months of age, which Werker and colleagues (e.g., 1981; Werker &

Tees, 1984a) attributed to their phonological irrelevance in the infants' language environment. Since the clicks are phonologically irrelevant in English, and fail to occur even as allophonic variants, the maintenance of discrimination for clicks calls for a modification of their argument, as outlined in the Introduction. However, based on our reasoning about perceptual assimilation, our findings are viewed as compatible with Werker's more general proposal of a developmental transition from prephonemic perception of speech sounds, which may entail a perceptual focus on either their auditory or their phonetic-articulatory properties (we favor the latter possibility: Best, 1984), to phonemic perception at around 10-12 months.

Of what use would the proposed transition to phonemic perception be for the infant's acquisition of the ambient language? The 10-12 month perceptual reorganization that Werker found closely parallels the universal milestones of beginning word comprehension and, for many infants, the first productions of words (e.g., Lenneberg, 1967; Stark, 1980; see also Ramsay, 1980, regarding language-related neuropsychological changes at this age). The prephonemic sensitivity of infants under 10-12 months of age for many nonnative contrasts is surely well-suited to their ability to learn whichever language surrounds them. However, as they become attuned to the ambient language and first begin to use words, phonemic perception should presumably aid their language acquisition. If phonemic perception entails assimilation of incoming sounds to the categories employed in the native language, then it may benefit the infant by sharpening the lines of structural organization within the phonological system of their language, and by helping to establish perceptual constancy among the acoustic variations of words pronounced in different contexts and by different speakers.

These benefits would presumably continue to aid efficient speech perception by adults, thus accounting for their continued difficulty with discriminating nonnative sounds that are assimilated to a single native phoneme category.

ACKNOWLEDGMENT

This research was supported by NIH grants HD-01994 and RR-05596 to Haskins Laboratories, NIH grant NS-24655 to the first author, and by a BRS grant awarded to the first author through Wesleyan University. The authors wish to thank the following people for their invaluable help in completing this research: Janet Werker for numerous useful discussions about speech perceptual development, helpful methodological advice, and the loan of her /ba/-/da/ stimuli; Andrea Levitt, Linda Polka, Bruno Repp, Michael Studdert-Kennedy, and three reviewers for thoughtful comments on earlier versions of the manuscript; Tshitshi Mbatha and Otty Nxumalo, for producing the Zulu click syllables from which our stimuli were chosen; Suzanne Margiano, Cynthia Nye, and Dianne Schrage for their assistance in collecting data from the infant subjects; our American and South African adult subjects; and most especially the parents who brought their infants to the laboratory to participate in the study.

REFERENCES

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academic.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 2*: Perception. New York: Academic Press.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development, 52,* 1135-1145.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B.M. Lester, & M. W. Yogman (Eds.), *Theory and research in behavioral pediatrics: Vol. 2.* New York: Plenum.

Best, C. T., MacKain, K. S., & Strange, W. (1982). *A cross-language study of categorical perception for semivowel and liquid contrasts.* Paper presented at 105th meeting of the Acoustical Society of America, April.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191-211.

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7,* 207-240.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Catford, J. C., & Ladefoged, P. (1968). Practical phonetic exercises. *UCLA Working Papers in Phonetics.* Los Angeles: UCLA Press.

Doke, C. M. (1926). The phonetics of the Zulu language. *Bantu Studies,* Special number.

Eilers, R. E., Gavin, W. & Oller, D. K. (1982). Cross-linguistic perception in infancy: Early effects of linguistic experience. *Journal of Child Language, 9,* 289-302.

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonetic perception in infancy: A cross-linguistic study. *Child Development, 50,* 14-18.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18,* 158-167.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Development of changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766-780.

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology. Vol. 8.* Berlin: Springer-Verlag.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science,* 171, 303-306.

Goldstein, L., & Browman, K. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14,* 339-342.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9,* 317-323.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch of the voicing cue. *Journal of the Acoustical Society of America, 47,* 613-617.

Horowitz, F. D. (1975). Visual attention, auditory stimulation, and language discrimination in infants. *Monographs of the Society for Research in Child Development, 39,* Serial No. 158.

Jusczyk, P. W. (1982). Auditory versus phonetic coding of speech signals during infancy. In J. Mehler, E. C. Walker, & M. Garrett (Eds.), *Perspectives on mental representation.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jusczyk, P. W. (1984). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), *Child phonology.* New York: Academic Press.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics.* Chicago: University of Chicago Press.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt, Brace & Jovanovich.

Ladefoged, P., & Traill, A. (1984). Linguistic phonetic descriptions of clicks. *Language, 60,* 1-20.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20,* 215-225.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74,* 431-461.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20,* 384-422.

MacKain, K. S. (1982). On explaining the role of experience on infants' speech discrimination. *Journal of Child Language, 9,* 527-542.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2,* 369-390.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge, England: Cambridge University Press.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on the perception of /ʃ/-/s/ distinction. *Perception & Psychophysics, 28,* 213-228.

Miller, C. L. (1983). Developmental changes in male-female voice classification by infants. *Infant Behavior and Development, 6,* 313-330.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331-340.

Muller, Z. (1965). *The world's living languages.* New York: Frederick Unger.

Nyembezi, C. L. S. (1972). *Learn Zulu.* Pietermaritzburg, South Africa: Shuter & Shooter Publishers.

Oller, D. K., & Eilers, R. E. (1983). Speech identification in Spanish- and English-learning 2-year-olds. *Journal of Speech and Hearing Research, 26,* 50-53.

Pisoni, D. B. (1973). Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253-260.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 297-314.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328-333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285-290.

Polka, K. (1987). *Perception of Persian uvular and velar stops by speakers of American English.* Paper presented at 113th meeting of the Acoustical Society of America, May.

Ramsay, D. S. (1980). Beginnings of bimanual handedness and speech in infants. *Infant Behavior and Development, 3, 67-77.*

Singh, S., & Black, J. W. (1966). Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America, 65 (Suppl. 1)* SS11.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1:* Production. New York: Academic Press.

Streeter, L. A. (1976). Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39-41.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38,* 579-590.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development, 47, 466-472.*

Weinreich, U. (1953) *Languages in contact.* New York: Linguistic Circle of New York.

Werker, J. F., Gilbert, J. V. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics,* 37, 35-44.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native sounds. *Canadian Journal of Psychology, 37,* 278-286.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Werker, J. F., & Tees, R. C. (1984b) Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Whalen, D., H. & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237,* 169-171.

Ziervogel, D., Louw, J. A., & Taljaard, P. C. (1976). *A handbook of the Zulu language.* Pretoria, South Africa: J. L. van Schaik.

FOOTNOTES

**Journal of Experimental Psychology: Human Perception and Performance,* in press.

†Also Wesleyan University

††Also University of Connecticut

†††Wesleyan University

[1]The perceptual process described here may be similar to the notion of "phonetic analogy" mentioned by Eilers et al. (1982) and also to the concept of "phonic interference" that has been used to describe the spoken errors made by learners of a second language (e.g., Weinreich, 1953). Note that the process proposed here is not the same phenomenon as "phonemic assimilation" in speech production whereby, for example, <pocketbook> is pronounced as though it were <pockepbook>.

[2]It should be noted that changes in the phoneme context of a given phoneme (i.e., changes in surrounding vowels and/or consonants) may affect the relation of specific phonetic properties to the phoneme category of interest. For example, /g/ is produced farther back in the context of /u/ than of /i/. Such context effects do have perceptual ramifications (e.g., Mann & Repp, 1980).

[3]The click at the apical place of articulation is similar to the "tsk" sound of disapproval. That at the lateral place is similar to the clicking sound that people sometimes make from one side of their tongue, along with a wink, when flirting or when showing approval or affectionate greeting. It may also occur with a wince, as a sign of regret or frustration, or when urging a horse along. The palatal click does not have a common nonverbal meaning in American society, as do the clicks at the other two places. It is similar, but not identical, to the "tongue cluck," a repetitive vocal play sound of some infants during the second half-year and into the preschool years, and the sound made to represent the clip-clop of a horse's slow gait.

[4]It should be noted that the "auditory experience" argument as presented by Aslin and Pisoni (1980) only mentions experience within a speech context. Of course, the underlying psychoacoustic assumptions of their view could easily be extended to predict that experience with clicks as nonspeech sounds should maintain sensitivity to them even in a speech context. However, if one were to predict systematic effects of nonspeech auditory experience upon developmental changes in speech perception, it would be difficult to decide which sorts of experience would be expected to have an effect and which nonnative contrasts would be affected in what particular directions. For example, would the sounds that the infant makes in early vocal development provide such auditory experience? Infant babbling includes not only clicks but also other non-English sounds such as trills and pharyngeal and uvular noises (Stark, 1980) that appear in contrasts that English-speaking adults and older infants find difficult to discriminate. See also Discussion, Experiment 1.

[5]These also participated in Experiment 3, as a comparison group for infants.

[6]The postshift trials were determined relative to the first time the infant fixated the slide, and thus heard at least one shift stimulus, after the habituation criterion had been reached. This definition is necessary because a stimulus shift has not occurred for the subject unless some audio shift stimuli have actually been presented. Some infants habituated to 0 fixation time, and then continued without any fixations during the first several slide presentations after the shift phase had been begun by computer (the shutter opens automatically after 1 s ITI, then closes automatically after 2 s of slide presentation without fixation, and continues to cycle through this way until the infant looks at the slide). For these infants, the first postshift trial with > 0 fixation was considered to be the de facto first postshift trial.

[7]We thank Michael Studdert-Kennedy for suggesting the interpretation discussed in this paragraph.

[8]We thank reviewer Kim Oller for reminding us of this important need.

Context Effects in Two-month-old Infants' Perception of Labio-dental/Interdental Fricative Contrasts*

Andrea Levitt,** Peter W. Jusczyk,† Janice Murray,†† and Guy Carden†††

We investigated two-month-old infants' perception of a subset of highly confusable English fricatives. In Experiment 1, infants discriminated modified natural tokens of the voiceless fricative pair [fa]/[θa], but only when the syllables included their frication noises. They also discriminated the voiced pair [va]/[da] both with and without fricative noises. These results parallel those found with adults by Carden, Levitt, Jusczyk, and Walley (1981). In Experiment 2, [f] and [θ] noises were appended to [a] and the same [f] noise was appended to the previously indiscriminable fricationless versions of [fa] and [θa]. Infants discriminated both pairs of stimuli, indicating 1) that the frication is a sufficient cue for [fa]/[θa] discrimination and 2) that it provides a context for discriminating the [f] and [θ] formant transitions. We conclude that infants' perception of labio-dental/interdental fricative contrasts shows evidence of context effects similar to those observed with adults.

Throughout much of its early history, infant speech perception research was directed at cataloging the kinds of contrasts that infants are capable of discriminating (e.g., Eimas, 1975; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Lasky, Syrdal-Lasky, & Klein, 1975; Miller & Eimas, 1979; Morse, 1972; Streeter, 1976; Trehub, 1976; Werker, Gilbert, Humphrey, & Tees, 1981). The accumulation of such knowledge about the extent of the infant's capacities was an important step forward in understanding speech perception in that it suggested that at least some of the underlying mechanisms are operative from birth. More recent investigations in the field have moved away from questions concerning the variety of contrasts that infants can discriminate to a consideration of the nature of the mechanisms themselves (e.g., Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983; Kuhl & Meltzoff, 1982; Miller & Eimas, 1983). Thus, there have been a number of investigations aimed at determining whether the mechanisms underlying the infant's discriminative capacities are speech-specific or general auditory ones (Eimas, 1974; Eimas & Miller, 1980; Jusczyk, Pisoni, Walley, & Murray, 1980). Such studies are informative in that they produce an indication of the range of acoustic signals to which the underlying mechanisms respond.

One recent discovery that holds promise with respect to furthering our understanding of perceptual mechanisms is the demonstration of context effects and trading relations in speech processing (e.g., Best, Morrongiello, & Robson, 1981; Oden & Massaro, 1978; Repp, 1982). Multiple cues figure in the categorization of speech along phonetic continua. There is evidence that the cues themselves enter into

trading relationships such that the presence of one cue can serve either to reinforce or offset the presence of another cue. For example, first formant onset frequency has been shown to enter into a trading relationship with voice onset time information in determining the locus of the voiced/voiceless category boundary in English (Summerfield, 1982; see also Pastore, Wielgus, & Szczesiul, 1984). The existence of such effects raises questions about the nature and origins of the mechanisms responsible for them.

There are several possible explanations for trading relations in speech perception. One possibility is that such context effects are the result of specific experience with a particular language. In effect, the listener learns how the cues are traded to ensure meaningful contrasts between words in his or her native language. In this case, some higher level decision rule might be imposed upon the perceptual output (e.g., see Oden & Massaro, 1978, for an account along these lines). The second possibility is that such trading relations are innately pre-wired either in the form of speech-specific or general auditory processing mechanisms. In fact, there is evidence that at least one type of context effect may be pre-wired, given that Miller and Eimas (1983) found evidence of context effects in infants as young as three months old. Specifically, they found evidence that changes in speech rate resulted in systematic shifts in the perception of cues to voicing and manner (stop/glide) distinctions.

Given the limited data available with infants, it is premature to conclude that trading relations and context effects are a consequence of the inherent organization of the underlying perceptual mechanisms. For while some such effects follow from the way in which the human auditory system is structured, others may have a different origin. In particular, some types of context effects may involve decision level processes that draw upon the listener's experience in producing and perceiving speech. One such possibility has been reported by Carden, Levitt, Jusczyk, and Walley (1981) in their investigation of the fricative contrast [f] (as in *fin*) and [θ] (as in *thin*). These researchers noted the existence of context effects in the perception of the place of articulation distinction that occurs between the syllables [fa] and [θa]. Specifically, the fricative noise and formant transition portions of these syllables interact in a way that gives rise to this distinction. Although the formant transition differences appear to function as a critical cue in distinguishing [fa] from [θa], they require a fricative context. Carden et al. demonstrated this in a series of experiments. First, they found that removing the fricative noise from [fa] and [θa] caused the resulting truncated syllables to be perceived as the same sound, [ba].[1] This result suggested that the fricative noise plays a critical role in signaling the [fa]-[θa] distinction. Yet further experimentation showed that it is not the information in the fricative noise *per se* that serves to distinguish [fa] from [θa]. In a separate unpublished study, we found that the addition of an identical fricative noise to both the truncated syllables for certain speakers was sufficient to reinstate the perception of these syllables as [fa] and [θa]. In other words, the fricative noise provided a necessary context for perceiving the minimal formant transition differences that distinguish [fa] and [θa] (cf. Harris, 1958).

To this point, these results are in line with the pattern typically observed for context effects. The suggestion that the effects might stem from decision level processes came as a result of another experiment (Carden et al., 1981) using synthetic speech sounds based on their natural tokens. By extending the range of the formant transition differences between [fa] and [θa], Carden et al. produced a synthetic speech continuum. Removing the fricative noise from this extended continuum resulted in a continuum that was perceived as extending from [ba] to [da]. Although composed of identical formant transitions, the two continua (one with and one without formant

transitions) had significantly different category boundary locations in a forced choice task. Thus, information about manner of articulation (i.e., whether the stimuli were stops or fricatives) affected subjects' decisions about place of articulation. This point was illustrated forcefully in a test condition in which subjects were presented with stimuli from the [ba]-[da] series but asked to label them as either [fa] or [θa]. The resulting category boundaries matched those of the original [fa]-[θa] series. A comparable pattern of results was observed for subjects presented with the [fa]-[θa] stimuli but asked to label them as [ba] or [da]. Thus, the context effects observed in labeling the formant transitions could occur even when the contextual cues were not physically appropriate. The results suggest that listeners employ different criteria in evaluating formant transition differences for stop and fricative distinctions. A possible factor in developing such criteria may well be the listener's own experience with the way in which such sounds are produced. If so, then this type of effect may be one that emerges only after considerable experience in producing and perceiving speech.

The present study was undertaken to examine the perception of fricative contrasts by young infants. More specifically, we sought answers to a number of questions, such as (1) do 2-month-olds discriminate contrasts between labiodental and interdental fricatives? (2) what kinds of information are critical for the infant's discrimination of such contrasts? and (3) do infants exhibit context effects for fricative contrasts similar to those observed by Carden et al. with adult subjects?

EXPERIMENT 1

That infants might experience difficulty in discriminating an [f]/[θ] contrast would not be surprising given that measures of speech intelligibility rank these sounds among the most highly confused in English (Miller & Nicely, 1955). The source for the confusability of these segments apparently lies in the similarity of their acoustic properties. Measurements made of the fricative portions of [f] and [θ] show that both are characterized by a rather uniform distribution of spectral energy (Klatt, 1986; Strevens, 1960), although the overall intensity may be somewhat lower in [θ] (Klatt, 1986). This suggests that some other portion of the signal, such as the formant transitions, may play the key role in distinguishing these segments. These observations are borne out by the results of perceptual studies suggesting that formant transition cues are sufficient in the productions of some speakers to distinguish [f] from [θ] (e.g., Harris, 1958; Heinz & Stevens, 1961).

There have been two previous studies that have examined the perception of the labiodental/interdental fricative distinction by infants. However, both studies (Eilers, Wilson, & Moore, 1977; Holmberg, Morgan, & Kuhl, 1977; see also Kuhl, 1980) examined infants who were 6 months old or older. By this age, infants have already had considerable experience listening to speech as well as some experience producing speech-like sounds. It is possible that even this limited experience is sufficient to allow infants to employ contextual cues in discriminating formant transition differences. Nevertheless, even at this age the infants gave some evidence of difficulty in discriminating these contrasts. For example, Holmberg et al. noted that their subjects required on average about twice as many trials to achieve a criterion for distinguishing [f]/[θ] as they did for a comparable contrast between [s]/[θ].[2] Given such considerations, it is reasonable to ask whether infants at a younger age and with considerably less experience are capable of discriminating a labiodental/interdental fricative contrast.

As in previous investigations, the present study focused on the ability of infants to discriminate between the voiceless fricative pair [fa] and [θa]. Moreover, following

the work of Carden et al. with adults, we also decided to examine the voiced counterpart to this distinction (i.e., [v] as in *vat* and [ð] as in *that*). The fricative noise portion of this latter pair tends to have a considerably lower amplitude than that of the voiceless pair. For this reason, the formant transition differences may take on even greater importance for discrimination. To explore further the contribution that formant transition differences might make to the infant's discrimination of fricative contrasts, two other types of contrasts were also examined. The stimuli involved in these contrasts were modified versions of the voiced and voiceless fricative pairs. The modification involved removing the fricative noise from the [fa]/[θa] and [va]/[ða] pairs, leaving the formant transitions and vocalic portions of the syllables intact. In line with the observations of Carden et al., removing the frication from the [fa]/[θa] pair resulted in a pair of syllables that both sounded like [ba] to adult listeners, whereas the modification of the [va]/[ða] pair produced syllables that preserved place of articulation information. This difference in subject response to the truncated versions of the voiceless and voiced fricative pairs is apparently due to the greater perceptual salience of the voiced formant transitions used to signal [va] and [ða].

To the extent that infants perceive the full and truncated versions of the syllables as do adults, they should discriminate both the full and truncated versions of the voiced [va]/[ða] pair, but only the full version of the voiceless [fa]/[θa] pair. Such a result would indicate that the infants not only discriminate fricative contrasts differing in place of articulation, but also may experience context effects in that the same set of formant transition differences are discriminable only in the presence of an appropriate fricative noise. On the other hand, if sensitivity to such context effects develops only after considerable experience in producing and listening to speech, and if the difference in the fricative noise spectrums is not a sufficient cue, then infants would not be expected to discriminate either version of the voiceless [fa]/[ða] pair, since the adult data suggest that formant transitions for these stimuli all fall within the same category. Under this second hypothesis, the predictions for the discriminability of the voiced pair are less clear. It may be that the formant transition differences are large enough, in which case both the full and truncated versions would be discriminated, in line with the adult data. Or, it may be that the formant transition differences are not sufficiently large, in which case neither pair would be discriminated.

Method

Procedure. Each infant was tested individually in a small laboratory room. The infant was seated in a reclining chair approximately .5 m away from a rear projection screen. An image of brightly colored flowers was projected on the screen for the entire test session. The projection screen was situated directly above a loudspeaker through which the auditory stimuli were presented. Each infant sucked on a blind nipple, held in place by an experimenter who wore headphones and listened to recorded music throughout the test session. A second experimenter in an adjacent room monitored the apparatus.

The experimental procedure was a modification of the high amplitude sucking (HAS) technique developed by Siqueland and DeLucia (1969). For each infant, the high-amplitude sucking criterion was established before the presentation of any stimuli. The criterion was set so as to produce a response rate of 15 to 30 sucks/min. Once a baseline rate of high-amplitude sucking was established, the presentation of stimuli was made contingent upon the rate of high-amplitude sucking. Since the stimuli ranged from 325 to 530 ms in duration with an interstimulus interval of approximately 500 ms, there was a maximum presentation rate of about one stimulus per sec. If the infant produced a burst of sucking responses with

interresponse times of less than one sec, then each response did not produce one presentation of the stimulus. Rather, the timing apparatus was reset so as to provide continuous auditory feedback for one sec after the last response of the sucking burst. Use of a programmable logic board ensured that all stimulus presentations were uninterrupted.

The criterion for satiation to the first stimulus was a decrement in sucking rate of 25% for two consecutive minutes compared to the rate of sucking in the immediately preceding minute. Once this criterion was met, the auditory stimulus was changed without interruption by switching channels on the tape recorder. For infants in the experimental conditions, the change was to an acoustically different stimulus. For infants in the control condition, the channels on the tape recorder were switched, but no acoustic change occurred because the same signal had been recorded on both channels. The postshift period lasted for four minutes. The infant's sensitivity to the change in auditory stimulation was inferred from comparisons during the postshift period.

Stimuli. Naturally produced syllables ([fa], [θa], [va], [ða]) were selected from one of the adult male talkers (P.N.) who produced the tokens used in the Carden et al. (1981) study. The tokens were recorded using an Ampex AG500 tape recorder. Each token was digitized at a 10 kHz sampling rate and low-pass filtered at 4.9 kHz (to prevent aliasing) using the Haskins Laboratories pulse code modulation (PCM) system (Cooper & Mattingly, 1969). As Carden et al.'s study demonstrated, the resulting syllables were clearly identifiable for adult listeners. The PCM system was also used to remove the post-transition vocalic portions of the [θa] and [ða] stimuli and replace them by the vocalic portions of the [fa] and [va] stimuli, respectively. This step was taken to ensure that the only difference between the stimuli lay in the frication and formant transitions. There was no indication of any perceptible spectral discontinuity in the resulting [θa] and [ða] tokens. The total duration of the [fa] and [θa] stimuli was 530 ms. The frication portions of these tokens had durations of 165 ms. This was achieved by removing the initial 20 ms of low amplitude frication from the [fa] to equate the fricative noise portions of the stimuli. The total duration of the voiced fricatives ([va] and [ða]) was 425 ms with the frication portion accounting for 100 ms. Truncated versions of the syllables were prepared using the PCM system to delete the frications from the digitized natural syllables cutting at the point of zero or near-zero amplitude nearest to the end of the frication. Here again, the vocalic portion was the same for each pair of stimuli. The resulting truncated (fricationless) stimuli are designated as [fa]-, [θa]-, [va]-, and [ða]-.[3] The output of the PCM system was then used to prepare the audio tapes employed in this experiment.

Design. Each infant in the study was seen for one session. Twelve subjects were assigned randomly to each of four groups and sixteen subjects were assigned randomly to a fifth (Control) group. Infants in Group I were tested for their ability to detect the [fa]/[θa] distinction, whereas subjects in Group II heard the truncated versions of these syllables (i.e., [fa]-/[θa]-). Groups III and IV were presented with contrasts involving [va]/[ða] and [va]-/[ða]-, respectively. The presentation order of the items in a given stimulus pair was counterbalanced across subjects for each group. Two each of the sixteen subjects in Group V were randomly assigned one of the eight stimuli for the entire test session.

Apparatus. A blind nipple was attached to a Grass PT5 volumetric pressure transducer, which in turn was connected to a Type DMP-4A Physiograph. A Schmitt trigger provided digital output of criterial high-amplitude sucking responses. Additional equipment included a Teac 3340 tape recorder, a Kenwood (KA-3500) amplifier, an ADS loudspeaker, a Grason-Stadler (Model 1200) programmable logic board, a power supply, two relays, a counter, and a Physiograph dc preamplifier. Each

criterial response activated the timer on the logic board for a 1-sec period or restarted the period. Auditory stimulation at a level of 72± dB (A) SPL (approximately 15 dB above the background noise level caused by the ventilation system) was available whenever the timer was in an active state. The use of the logic board to monitor the auditory signals on the tape ensured that the timer was never activated in themiddle of a stimulus.

Subjects. The subjects were 64 infants, 36 males and 28 females. Mean age was 9.5 weeks (range: 6 to 12 weeks). In order to obtain complete data on 64 subjects, it was necessary to test 136. Subjects were excluded for the following reasons: crying (42%), falling asleep (32%), ceasing to suck during the course of the experiment (3%), failure to meet the habituation criterion within 24 minutes (9%), failure to acquire the the response (3%), equipment failure and experimenter error (4%), and miscellaneous (hiccups, etc.)(7%).[4]

Results

For purposes of statistical comparison, an examination was made of each subject's rate of sucking during five intervals: baseline minute, third minute before shift, average of the last two minutes before shift, average of the first two minutes after shift and average of all four minutes after shift. Difference scores were then calculated for each subject for each of the following comparisons: (1) acquisition of the sucking response: third minute before shift minus baseline; (2) satiation: third minute before shift minus the average of the last two minutes before shift; (3) release from satiation: average of the first two minutes after shift minus the average of the last two minutes before shift; (4) release from satiation for the full four minutes: average of the four minutes after shift minus the last two minutes before shift.

TABLE 1
Mean Change in Response Rate after Shift

		RELEASE FROM SATIATION (MINUTES AFTER SHIFT)	
	Stimulus Pair	*First 2*	*Full 4*
Group I	fa/θa	7.04*	7.40*
Group II	fa-/θa-	.71	.42
Group III	va/ða	5.71*	4.15*
Group IV	va-/ða-	9.96*	6.56*
Group V	CONTROL	-1.47	-1.59

*Indicates a reliable difference ($p<.05$ or better) according to randomization tests when compared against performance in the Control Session.

Subjects in all conditions acquired the high-amplitude response and satiated to the first stimulus prior to shift. An indication of the mean change in response rate during the period following shift for each of the five groups is provided in Table 1. Randomization tests for independent samples (Siegel, 1956) were used to assess postshift performance of each of the experimental groups in comparison to the control group for the first two minute and full four minute release from satiation measures. Because the pattern of significant results ($p<.05$ or better, one-tailed) was

the same for both the two- and four-minute postshift periods, the subsequent discussion will not distinguish between them (see Table 1). The statistical analysis revealed that Groups I, III, and IV ([fa]/[θa], [va]/[ða], and [va]-/[ða]-) displayed significant increases in sucking relative to the control group. Group II ([fa]-/[θa]-) did not differ from the control group. Thus, the infants behaved in accordance with the adultlike pattern in that they discriminated both of the voiced fricative contrasts, but only the voiceless fricative contrast in which frication noise was present.

Discussion

The present results indicate that infants as young as two months old are capable of discriminating place of articulation contrasts in voiced and voiceless fricative pairs. To the best of our knowledge, the present study is the first to show discrimination of the voiced pair, [va]/[ða], by infants. The finding that infants also discriminated the voiceless pair is consistent with the finding by Holmberg et al. that 6-month-old infants can discriminate an [f]/[θ] contrast.

The present study also explored the role of formant transitions in the infant's perception of fricative contrasts. Formant transition differences do appear to provide a sufficient basis for the infant's discrimination of the voiced fricative contrast as evidenced by the fact that both the versions with (i.e., [va]/[ða]) and without (i.e., [va]-/[ða]-) the appropriate frication noise are discriminated. A different picture is presented by the results with the voiceless fricatives. Formant transition differences for these items proved to provide sufficient basis for discrimination only when accompanied by the appropriate frication noise. One possible reason for the infants' failure to discriminate the [fa]-/[θa]- pair is that the fricative noises provide the distinctive cues for signaling the contrast. This would not be surprising given that the overall amplitude of fricative noise for voiceless fricatives is considerably higher than for their voiced counterparts. Moreover, as noted earlier in the discussion of Carden et al.'s results, the formant transition differences seem to be less distinctive for voiceless than for voiced fricatives. Nevertheless, there is a second possibility that needs to be considered, viz., that the fricative noise provides a necessary context for discriminating the formant transition differences. By this latter view, it is not that there are distinctive cues inherent in the frication noises of [fa] and [θa], but rather the processing of the formant transitions as cues to a place of articulation difference depends upon their being perceived as part of an articulatory gesture relating to fricative production.[5] Indeed, the results of Harris (1958), demonstrating that frication noises from [fa] and [θa] syllables can be interchanged with no apparent change in their identification, support the view that the noise may provide only the necessary context as opposed to distinctive information.[6] In any event, the present results do not allow us to distinguish between these two alternative explanations for the infants' failure to discriminate the [fa]-/[θa]- contrast. For this reason, we decided to undertake a more systematic investigation of the role that fricative noises play in infants' discrimination of voiceless fricative contrasts.

EXPERIMENT 2

The notion that [f] and [θ] noise may primarily play a contextual role in the perception of voiceless fricative contrasts stems from the findings that interchanging these fricative noises apparently does not change the perceived identity of the resulting sounds for adult listeners (Harris, 1958). By extension, one would expect that appending the same fricative noise (e.g., one appropriate to [fa]) to fricationless versions of [fa] and [θa] syllables would have much the same effect, viz.,

that these syllables would be heard as [fa] and [θa], respectively. Thus, if fricative noise merely serves as a necessary context for discriminating the formant transition differences between [fa] and [θa], then the addition of a common fricative noise to the [fa]-/[θa]- tokens should allow infants to discriminate them. On the other hand, if the role of the fricative noise is to provide distinctive cues to the identity of [fa] and [θa], then the presence of [fa] vs. [θa] frication differences may be a sufficient basis for discrimination, even in the absence of any accompanying formant transition differences. The present experiment was designed to test both of these possibilities. Hence, one group of infants heard a contrast between [fa]- and [θa]- tokens to which a common [f]-fricative noise had been appended. A second group of infants was presented with a contrast between items consisting of the frication portions of [fa] and [θa] plus the vowel with the distinctive formant transitions removed.

Method

Procedure. The HAS procedure was employed as described in the previous experiment.

Stimuli. The [fa] and [θa] tokens employed in Experiment 1 were modified for use in the second experiment. The PCM system was used to make the necessary modifications. One pair of stimuli consisted of the original [fa] stimulus plus a hybrid stimulus produced by taking the [θa] token, removing its frication by cutting at the point of zero amplitude nearest to the end of fricative noise, and substituting the comparable frication noise from [fa] token. Since the post transition vocalic portions of the original [fa] and [θa] tokens were identical, this new token, "Fn+θa-" ([f] noise + fricationless [θa]), differed from the [fa] stimulus only in its formant transitions, which were appropriate for [θa]. Its overall duration was 530 ms, the same as the [fa] token. Note that there were no obvious acoustic discontinuities in this sound.

The second pair of items was produced by using the PCM system to remove the formant transitions of the vocalic portion of the [fa] token used in Experiment 1. The frication noises from [fa] and [θa], obtained by cutting the stimuli just prior to the first pitch pulse, were appended to the common vocalic portion. The resulting tokens, designated as "Fn+a" ([f] noise + [a]) and "θn+a" ([θ] noise + [a]), had overall durations of 475 ms.[7] The output of the PCM system was then used to prepare the audio tapes employed in this experiment.

Design. Each infant in the study was seen for one session. Twelve subjects were assigned randomly to each of two experimental groups and an additional twelve subjects were assigned randomly to a control group. Infants in one group were tested for their ability to detect a distinction between [fa] and Fn+θa-, while infants in a second group were tested on the Fn+a/θn+a contrast. The presentation order of the items in a given stimulus pair was counterbalanced across subjects for each group. Three each of the twelve control subjects were assigned at random to one of the four stimuli for the entire test session.

Apparatus. The same equipment was employed as described for the previous experiment.

Subjects. The subjects were 36 infants, 18 males and 18 females. Mean age was 10.1 weeks (range: 7.6 to 11.6 weeks). In order to obtain complete data on 36 subjects, it was necessary to test 75. Subjects were excluded for the following reasons: crying (59%), falling asleep (28%), equipment failure (5%), failure to acquire the response (5%) and miscellaneous (hiccups, etc.) (3%).

Results

Difference scores were calculated for each subject as per Experiment 1 for (1) acquisition of the sucking response; (2) satiation to the preshift stimulus; (3) release from satiation during the first two postshift minutes; and (4) release from satiation for the full four postshift minutes. As in the previous experiment, subjects in all groups acquired the conditioned response and satiated to the preshift stimulus. An indication of the mean change in response rate during the postshift period is displayed in Table 2. Randomization tests for independent samples were again employed to analyze postshift performance for both the two- and full-four-minute periods. The pattern of significant results ($p<.025$ or better, one-tailed)[8] was identical for both the two- and full-four-minute periods. Both of the experimental groups (i.e., [fa]/Fn+ea- and Fn+a/en+a) exhibited significant increases in postshift sucking relative to the controls. Thus, fricative noise evidently contributes to the infant's perception of fricative contrasts in two ways -- both as a source of distinctive information and as a context for discriminating formant transition differences.

TABLE 2
Mean Change in Response Rate after Shift

		RELEASE FROM SATIATION (MINUTES AFTER SHIFT)	
	Stimulus Pair	*First 2*	*Full 4*
Group I	en+a/Fn+a	6.29*	6.52*
Group II	Fn+ea-/Fa	11.29*	10.75*
Group III	CONTROL	-4.42	-4.29

*Indicates a reliable difference ($p<.025$ or better) according to randomization tests when compared against performance in the Control Session.

Discussion

The present experiment investigated the role that fricative noise plays in infants' discrimination of fricative contrasts. Specifically, does the importance of the fricative noise lie in distinctive cues that it embodies or does it merely provide an appropriate context for formant transition cues in signaling a place of articulation contrast between fricatives? The somewhat surprising answer seems to be that it does both. Consider first the notion that distinctive cues are inherent in the frication. The present experiment demonstrated that the addition of only the frication portion of [fa] and [ea] to the same vocalic segment ([a]) resulted in discriminably different tokens for infants. Given this result and the demonstration from the previous experiment that infants did not discriminate truncated fricative syllables lacking the appropriate frications, one might be tempted to conclude that the distinctive fricative noises are both the necessary and sufficient cues for infants. However, this conclusion must be rejected given the results for the second experimental group in the present study. Infants were able to discriminate the [fa]/Fn+ea- contrast despite the fact that both items included an identical frication portion. Instead, the two members of this pair differed only in their formant transitions.[9] This latter result parallels the sorts of context effects observed by Carden et al. (1981) using synthetic stimuli with adult subjects. Thus, under the proper circumstances, it appears that

infants are able to utilize either formant transition or fricative noise differences to signal the [fa]/[θa] contrast.

Further statistical support for this conclusion comes from an additional analysis that we conducted comparing the performance of some of the groups across the two experiments. Note that across these experiments all crucial combinations of differences in fricative spectrum and formant transitions were tested. Thus, Group I in Experiment 1 (fa/θa) received stimuli that differed in both the frication spectrum and formant transitions, whereas Group II (fa-/θa-) received stimuli that differed on formant transitions without any accompanying frication. In Experiment 2, Group I (θn+a/Fn+a) received stimuli that differed only in their frication portion, but without any formant transition and Group II (Fn+θa-/fa) received stimuli that differed in formant transitions but included the same frication context. A Kruskal-Wallis one-way ANOVA indicated the main effect for groups was significant, $2(3)=7.84$, $p<.05$. Post hoc analyses conducted with randomization tests for independent samples revealed that the relatively poor discrimination response of the fa-/θa- group was largely responsible for the effect. Specifically, both Fn+a/θn+a and the Fn+θa-/fa groups had significantly higher postshift sucking scores, $t(22)=3.17$, $p<.005$ and $t(22)=1.72$, $p<.05$, respectively, than the fa-/θa- group. Similarly, the postshift performance of the fa/θa group was in the same direction, although marginal, $t(22)=1.69$, $p<.06$. None of the other group comparisons remotely approached significance.

GENERAL DISCUSSION

The present study was designed to examine several aspects of young infants' perception of place of articulation for fricatives. In particular, it asked whether the [fa]/[θa] and [va]/[ða] contrasts were discriminable for two month olds given the reports that such contrasts were relatively difficult for older infants (Eilers et al., 1977; Holmberg et al., 1977). In this matter, the results were unambiguous in indicating that such contrasts are discriminable for two month olds.

Having established that infants have the capacity to discriminate such contrasts at an early age, we sought to determine the nature of the information that infants were responding to. The first experiment suggested that formant transition differences alone were not a sufficient basis to account for infants' discrimination of the [fa]/[ða] contrast because there was no evidence of discrimination in the absence of an appropriate fricative noise context. Indeed, this parallels the results found with adults (Carden et al., 1981). This led to an investigation of the role that fricative noise plays in the discrimination of the contrast. Despite the fact that most acoustic analyses reveal great similarities in the spectral characteristics of the frication portions of [fa] and [θa], there must be some distinctive components because infants were able to discriminate the frications in the absence of any other distinctive cues.

At the very least, the frication portion of the signal provides a necessary context for discriminating the kinds of formant transition differences found in natural utterances of [fa] and [θa]. This was demonstrated by the fact that the addition of a common fricative noise to a previously indiscriminable formant transition contrast served to render it discriminable. This result parallels the kinds of context effects observed in connection with adults' perception of fricative contrasts (Carden et al., 1981). It indicates that the context effects themselves do not depend on a long apprenticeship in producing and listening to speech. Rather, the source of these effects appears to be a consequence of the inherent organization of the underlying perceptual mechanisms.

The present findings point to a number of potentially useful directions for research towards understanding the mechanisms responsible for context effects. First, recall that the suggestion that some context effects may have a basis in linguistic experience stemmed from Carden et al.'s observation that simply instructing subjects to use a fricative or stop context was sufficient to produce category boundary shifts along synthetic speech continua. Given the present results, it may be that what is acquired with linguistic experience is not the different boundary locations for stop and fricative continua, but the ability to infer the necessary context when the cues are not physically present. Consequently, one would anticipate that infants would display different category boundary locations for stop and fricative continua produced by varying formant transitions. If so, this would be further evidence that the context effects themselves are inherent in the underlying perceptual mechanisms. We are currently investigating this possibility and undertaking further studies to see whether infants can be induced to shift boundaries in the absence of the appropriate physical context.

A further direction for research is to attempt to establish whether the underlying perceptual mechanisms are general auditory ones or whether they are specific to speech processing. In the case of previous reports of context effects in the processing of speech contrasts (i.e., Miller & Eimas, 1983), there was evidence of comparable perceptual boundary shifts for certain nonspeech stimuli (Jusczyk et al., 1983). It would be useful to know whether effects comparable to the ones observed in the present study occur in the infant's perception of nonspeech sounds.

In summary, the present study demonstrates that infants as young as two months old do have the capacity to discriminate place of articulation differences in labiodental and interdental fricatives. The results also suggest that the presence of an appropriate fricative noise context is a critical factor in the way in which the distinctive formant transition cues to such contrasts are perceived.

ACKNOWLEDGMENT

This research was supported in part by NICHD Grant HD01994 and BRS Grant 05596 to Haskins Laboratories and in part by grants from NSERC (A-0282) and NICHD (HD 15795) to PWJ. We are grateful to Christopher Murphy and Anne Ferguson for their assistance in running subjects. We also thank Catherine Best, Alvin Liberman, Susan Nittrouer, and Douglas Whalen for helpful comments on previous versions of this manuscript. Portions of this research were reported at the biennial meeting of the Society for Research in Child Development in San Francisco, March 17, 1979.

REFERENCES

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191-211.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics, 29,* 26-36.

Cooper, F. S., & Mattingly, I. G. (1969). A computer-controlled PCM system for the investigation of dichotic speech perception. *Journal of the Acoustical Society of America, 46,* 115 (A).

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America, 75,* 897-907.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766-780.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16,* 513-521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics, 18,* 341-347.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209,* 1140-1141.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech, 1,* 1-7.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America, 33,* 589-596.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). *Speech perception in early infancy: Discrimination of fricative consonants.* Paper presented at the 94th Meeting of the Acoustical Society of America, Miami Beach, Florida, December 16.

Jusczyk, P. W. (1981). Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech.* Hillsdale, NJ: Erlbaum.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A., & Myers, M. (1983). Durational context effects in the processing of nonspeech sounds by infants. *Science, 222,* 175-176.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America, 67,* 262-270.

Klatt, D. H. (1986). Problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum.

Kuhl, P. K. (1980). Perceptual constancy for speech sound categories in early infancy. In G. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology, Vol. 2. *Perception.* New York: Academic Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138-1141.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. (1975). VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology, 20,* 213-225.

Miller, G. A., & Nicely, P. E. (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 338-353.

Miller, J. L., & Eimas, P. D. (1979). Organization in infant speech perception. *Canadian Journal of Psychology, 33,* 353-367.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135-165.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology, 14,* 477-492.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172-191.

Pastore, R. E., Wielgus, V. G . & Szczesiul, R. (1984). F1-cutback interactions in the perception of voicing contrasts. *Journal of the Acoustical Society of America, 76,* 28. (A).

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81-110.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Siqueland, E. R., & DeLucia, C. A. (1969). Visual reinforcement of non-nutritive sucking in human infants. *Science, 165,* 1144-1146.

Streeter, L. A. (1976). Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature, 259,* 39-41.

Strevens, P. (1960). Spectra of fricative noise in human speech. *Language and Speech, 3,* 32-49.

Summerfield, A. Q. (1982). Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops. *Journal of the Acoustical Society of America, 72,* 51-61.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47,* 466-472.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

FOOTNOTES

Journal of Experimental Psychology: Human Perception and Performance, in press

**Also Wellesley College

†Department of Psychology, University of Oregon and Laboratoire de Science Cognitive et Psycholinguistique. C.N.R.S. & E.H.S.S., Paris

††Department of Psychology, University of Auckland

†††University of British Columbia

[1]Although there was some individual variation in the way in which truncated versions of tokens from different speakers were identified by listeners in Carden et al.'s study, they all showed the same pattern.

[2]Note that Eilers et al. (1977) reported that 6- and 12-month old infants were unable to discriminate a [fa]/[θa] contrast. Jusczyk (1981) suggested that one possible reason for this result was that the tokens used by Eilers et al. were correctly identified by adults only 70% and 60% of the time for [fa] and [θa] respectively, whereas P. N.'s [fa] was correctly identified 92% and his [θa] 69% of the time. Some explanation along these lines seems reasonable, given the success reported by Holmberg et al. with infants in this age range.

[3]In the case of talker P.N., when the [fa]- and [θa]- tokens were identified as stop consonants, they were labeled as "ba" 99% and 91% of the time, respectively.

[4]The dropout rate for this experiment is typical for experiments with infant subjects. Furthermore, subjects were eliminated only during the preshift portion of the experiment and not after, so that it might be more accurate to say that the infants who completed the preshift portion of the test were assigned randomly to the experimental groups.

[5]One possible source for such an effect may lie in peripheral auditory processing. Delgutte and Kiang (1984) have argued, based on single unit auditory nerve recordings, that fricative noise modifies the coding of the following consonant transitions. On the other hand, this explanation, in terms of an auditory contrast effect, does not easily account for Carden et al.'s (1981) results where subjects showed a boundary shift on a transition continuum when they were merely instructed to assign fricative labels to the stop (i.e., fricationless) series.

[6]Our preliminary studies confirm the results of Harris (1958), but also indicate that there is a great deal of speaker variation.

[7]The neutral [a] context was included because it did not appear that infants would suck to hear the fricative noise portions in isolation. We do not wish to make any claim for the naturalness of these tokens or for the absence of discontinuity between the frication and vocalic portions of these stimuli, although no such discontinuity was perceptible to us. Suffice it to say that any discontinuities would be the same for both members of the pair.

[8]One-tailed tests are typically used in this type of experimental study with infant subjects, since the prediction is always that the infants in the test groups will show greater release from satiation than those in the control group and not less. Nonetheless, these results were also significant in a two-tailed test ($p<.05$ or better).

[9]Informal testing with adults showed that the particular token we used of Fn+θa- was not strongly perceived as an interdental. Thus, it is all the more striking that infants made this discrimination and lends strong support for interpreting such results as a context effect.

The Phoneme as a Perceptuomotor Structure*

Michael Studdert-Kennedy†

Studies of speech and writing face a paradox: the discrete units of the written representation of an utterance cannot be isolated in its quasi-continuous articulatory and acoustic structure. We may resolve the paradox by positing that units of writing (ideographs, syllabic signs, alphabetic letters) are symbols for discrete, perceptuomotor, neural control structures, normally engaged in speaking and listening. Focusing on the phoneme, for which an alphabetic letter is a symbol, the paper traces its emergence in a child's speech through several stages: hemispheric specialization for speech perception at birth, early discriminative capacity followed by gradual loss of the capacity to discriminate among speech sounds not used in the surrounding language, babbling, and first words. The word, a unit of meaning that mediates the child's entry into language, is viewed as an articulatory routine, a sequence of a few variable gestures of lips, jaw, tongue, velum, and larynx, and their acoustic correlates. Under pressure from an increasing vocabulary, recurrent patterns of sound and gesture crystallize into encapsulated phonemic control units. Once a full repertoire of phonemes has emerged, usually around the middle of the third year, an explosive growth of vocabulary begins, and the child is soon ready, at least in principle, for the metalinguistic task of learning to read.

> Ever since I...started to read...there has never been a line that I didn't hear. As my eyes followed the sentence, a voice was saying it silently to me. It isn't my mother's voice, or the voice of any person I can identify, certainly not my own. It is human, and it is inwardly that I listen to it. Eudora Welty (1983, p. 12)

INTRODUCTION

Any discussion of the relation between speech and writing faces a paradox: the most wide-spread and efficient system of writing, the alphabet, exploits a unit of speech, the phoneme, for the physical reality of which we have no evidence. To be sure, we have evidence of its psychological reality. But, ironically, that evidence depends on the alphabet itself. How are we to escape from this circle?

First, let me elaborate the terms of the paradox. Since the earliest spectrographic, cineradiographic, and electromyographic studies, we have known that neither the articulatory nor the acoustic flow of speech can be divided into a sequence of segments corresponding to the invariant segments of linguistic description. Whether the segments are words, morphs, syllables, phones, or features, the case is the same. The reason for this is simply that we do not normally speak phoneme by phoneme, syllable by syllable, or even word by word. At any instant, our articulators are executing a complex interleaved pattern of movements of which the spatio-temporal coordinates reflect the influence of several neighboring segments. The typical result is that any isolable articulatory or acoustic segment arises as a vector of forces from

more than one linguistic segment, while any particular linguistic segment distributes its forces over several articulatory and acoustic segments. This lack of isomorphism between articulatory-acoustic and linguistic structure is the central unsolved problem of speech research (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Pisoni, 1985). Its continued recalcitrance is reflected in the fact that (apart from a variety of technologically ingenious, but limited and brute force solutions) we are little closer to automatic speech recognition today than we were thirty years ago (Levinson & Liberman, 1981).

What then is the evidence for the psychological reality of linguistic segments? (I confine my discussion to the phoneme, although most of what follows would apply *mutatis mutandis* to all other levels of description.) First and foremost is the alphabet itself. Superficially, we might take the alphabet (or any other writing system) to be a system of movement notation analogous to those used by ethologists to describe, say, the mating behavior of Tasmanian devils (Golani, 1981). The difference lies in their modes of validation. The ethologist's units may or may not correspond to motor control structures in the devil's behavior; the units are sufficiently validated, if they lend order and insight to the ethologist's understanding of that behavior. By contrast, the alphabet (like music and dance notation) is validated by the fact that it serves not only to notate, but to control behavior: we both write and read. Surely, we could not do so with such ease, if alphabetic symbols did not correspond to units of perceptuomotor control. A writing system constructed from arbitrary units—phonemes and a half, quarter words—would be of limited utility. We infer then that lexical items (words, morphemes) are stored as sequences of abstract perceptuomotor units (phonemes) for which letters of an alphabet are symbols.

If this is so, those who finger the phoneme as a fictitious unit imposed on speech by linguists because they know the alphabet (e.g., Warren, 1976) have it backwards. Historically, the possibility of the alphabet was discovered, not invented. Just as the bicycle was a discovery of locomotor possibilities implicit in the cyclical motions of walking and running, so the alphabet was a discovery of linguistic possibilities implicit in patterns of speaking.

Of course, we do have other important sources of evidence that confirm the psychological reality of the phoneme: errors of perception (e.g., Browman, 1980) and production[1] (e.g., MacKay, Chapter 18; Shattuck-Hufnagel, 1983), backward talking (Cowan, Leavitt, Massaro, & Kent, 1982), aphasic deficit (e.g., Blumstein, 1981). But we can only collect such data because we have the metalinguistic awareness and notational system to record them. Illiterates may make speech errors (MacKay, 1970), and oral cultures certainly practice alliteration and rhyme in their poetry. But, like the illiterate child who relishes "Hickory dickory dock," they probably do not know what they are doing (cf. Morais, Cary, Alegria, & Bertelson, 1981). Thus, the data that confirm our inferences from the alphabet rest squarely on the alphabet itself.

The paradox I have outlined might be resolved, if we could conceptualize the relation between a letter of the alphabet (or a word) and the behavior that it symbolizes. Just how difficult this will be becomes apparent if we compare the information conveyed by a spoken word with the information conveyed by its written counterpart. An experimenter may ask a willing subject either to repeat a spoken word or to read aloud its written form. The subject's utterances in the two cases will be indiscriminable, but the information that controlled the utterances will have been quite different. The distinction, due to Carello, Turvey, Kugler, and Shaw (1984) (see also Turvey & Kugler, 1984) is between information that *specifies* and information that *indicates* or *instructs*. The information in a spoken word is not arbitrary: its acoustic structure is a lawful consequence of the articulatory gestures

that shape it. In other words, its acoustic structure is *specific* to those gestures, so that the prepared listener can follow the specifications to organize his own articulation and reproduce the utterance. Of course, we do not need the full specification of an utterance, in all its phonetic detail, in order to perceive it correctly, as those who know a foreign language, yet speak it with an accent, demonstrate: capturing all the details calls for a subtle process of perceptuomotor attunement. But it is evident that the specification does suffice for accurate reproduction, given adequate perceptuomotor skill, both in the child who slowly comes to master a surrounding dialect and in the trained phonetician who precisely mimics that dialect.

By contrast, the form of a written word is an arbitrary convention, a string of symbols that *indicate* to a reader what he is to do, but do not tell him how to do it. What is important here is that indicational information cannot control action in the absence of information specific to the act to be performed. That is why we may find it easier to imitate the stroke of a tennis coach than to implement his verbal instructions. Similarly, we can only pronounce a written word if we have information specifying the correspondences between the symbol string and the motor control structures that must be engaged for speaking. These are the correspondences that an illiterate has not discovered.

The question now is simply this: what is the relation between a discrete symbol and the continuous motor behavior that it controls? If a written symbol does indeed stand for a motor control structure, as argued above, we may put the question in a slightly more concrete form: What is the relation between a discrete motor control structure and the complex pattern of movements that it generates? The answer will certainly not come in short order. But perhaps we can clarify the question, and gain insight into possible lines of answers by examining how units of perceptuomotor control emerge, as a child begins to speak its first language.

Basic to this development is the child's capacity to imitate, that is, to reproduce utterances functionally equivalent to those of the adults around it. We have claimed above that an utterance specifies the articulation necessary to reproduce it. But until we spell out what specification entails, the claim amounts to little more than the observation that people can repeat the words they hear. At least three questions must be answered, if we are to put flesh on the bones.

First is the question of how a listener (or, in lipreading, a viewer) transduces a pattern of sound (or light) into a matching pattern of muscular controls, sufficient to reproduce the modeled event. We can say very little here other than that the acoustic/optic pattern must induce a neural structure isomorphic with itself. The pattern must be abstract in the sense that it no longer carries the marks of its sensory channel, but concrete in that it specifies (perhaps quite loosely, as we shall see below) the muscular systems to be engaged: no one attempts to reproduce a spoken utterance with his feet. The perceptuomotor structure is therefore specific to the speech system. Perhaps it is worth remarking that, in the matter of transduction, the puzzle of imitation seems to be a special case of the general puzzle of how an animal modulates its actions to fit the world it perceives.

The second question raised by imitation concerns the units into which the modeled action is parsed. Research in speech perception has been preoccupied with units of linguistic analysis: features and phonemes. These, as normally defined, are abstract units, unsuited to an account of imitation, because, whatever their ultimate function in the adult speaker, they do not correspond to primitives of motor control that a child might engage to imitate an utterance in a language that it does not yet know. The human vocal apparatus comprises several discrete, partially independent articulators (lips, jaw, tongue, velum, larynx) by which energy from the respiratory system is modulated. The perceptual units of imitation must therefore be structures that specify functional units of motor control, corresponding to actions of the

articulators. Isolation of these units is a central task for future research. We will come back to this matter below.

A third issue for the study of speech imitation is the notorious many-to-one relation between articulation and the acoustic signal (Porter, Chapter 5). Speakers who normally raise and then lower their jaws in producing, say, the word, "Be!", may execute acoustically identical utterances with pipes clenched between their teeth. The rounded English vowel of, say, *coot* may be produced either with protruded lips and the tongue humped just in front of the velum, or with spread lips, the tongue further backed and the larynx lowered. Even more bizarre articulations are discovered by children, born without tongue blade and tip, who nonetheless achieve a surprisingly normal phonetic repertoire (MacKain, 1983). Thus, the claim that an utterance specifies its articulation cannot mean that it specifies precisely which articulators are to be engaged, and when. Rather, it must mean that the utterance specifies a range of functionally equivalent articulatory actions. Of course, functional (or motor) equivalence is not peculiar to speech and may be observed in animals as lowly as the mouse (Fentress, 1981, 1983; Golani, 1981). Solution of the problem is a pressing issue in general research on motor control. For speech (and for other forms of vocal imitation, in songbirds and marine mammals) we have an added twist: the arbiter of equivalence is not some effect on the external world—seizing prey, peeling fruit, closing a door—but a listener's judgment.

EARLY PERCEPTUAL DEVELOPMENT

With all this in mind, let us turn to the infant. Perceptually, speech already has a unique status for the infant within a few hours or days of birth. Neonates discriminate speech from non-speech (Alegria & Noirot, 1982), and, perhaps as a result of intrauterine stimulation, prefer their mothers' voices to strangers' (DeCasper & Fifer, 1980). Studies of infants from one to six months of age, using a variety of habituation and conditioning techniques, have shown that infants can discriminate virtually any speech sound contrast on which they are tested, including contrasts not used in the surrounding language (see Eimas, 1985, for review). However, similar results from lower animals (chinchillas, macaques) indicate that infants are here drawing on capacities of the general mammalian auditory system (see Kuhl, 1986, for review).

Dissociation of left and right sides of the brain for speech and nonspeech sounds, respectively, measured by relative amplitude of auditory evoked response over left and right temporal lobes, may be detected within days of birth (Molfese, 1977). Left and right hemisphere short-term memories for syllables and musical chords, respectively, measured by habituation and dishabituation of the cardiac orienting response to change, or lack of change, in dichotic stimulation, are developing by the third month (Best, Hoffman, & Glanville, 1982). These and other similar results (see Best et al., 1982, and Studdert-Kennedy, 1986, for review) are important, because many descriptive and experimental studies have established that speech perceptuomotor capacity is vested in the left cerebral hemisphere of more than 90% of normal adults.

At the same time, we should not read these results as evidence of "hard wiring." At this stage of development not even the modality of language is fixed. If an infant is born deaf, it will learn to sign no less readily than its hearing peers learn to speak. Recent studies of "aphasia" in native American Sign Language signers show striking parallels in forms of breakdown between signers and speakers with similar left hemisphere lesions (Bellugi, Poizner, & Klima, 1983). Thus, the neural substrate is shaped by environmental contingencies, and the left hemisphere, despite its predisposition for speech, may be usurped by sign (Neville, 1980, 1985; Neville, Kutas,

& Schmidt, 1982). Given the diversity of human languages to which an infant may become attuned, such a process of epigenetic development is hardly surprising.

EARLY MOTOR DEVELOPMENT

The development of motor capacity over the first year of life may be divided into a period before babbling (roughly, 0-6 months) and a period of babbling (7-12 months) (Oller, 1980). At birth, the larynx is set relatively high in the vocal tract, so that the tongue fills most of the oral cavity, limiting tongue movement and therefore both the possible points of intraoral constriction, or closure, and the spectral range of possible vocalic sounds. Accordingly, early sounds tend to be neutral, vowel-like phonations, often nasalized (produced with lowered velum), with little variation in degree or placement of oral constriction. As the larynx lowers, the variety of nonreflexive, nondistress sounds increases. By the second trimester, sounds include labial trills ("raspberries"), squeals, and primitive syllabic patterns, formed by a consonant-like closure followed by a vowel-like resonance. These syllabic patterns lack the precise timing of closure, release, and opening characteristic of mature syllables.

In fact, the onset of true or canonical babbling (often a quite sudden event around the seventh month) is marked by the emergence of syllables with the timing pattern (including closing to opening ratio), typical of natural languages (Oller, 1986). In the early months, syllables tend to be reduplicated (e.g., [bababa], [mamama], [dadada]); these give way in later months to sequences in which both consonant and vowel vary. Phonetic descriptions of babbled consonants (e.g., predominance of stops, glides, nasals, scarcity of fricatives, liquids, consonant clusters) tend to be similar across many language environments, including that of the deaf infant (Locke, 1983). We may therefore view these preferences as largely determined by universal anatomical, physiological, and aerodynamic constraints on vocal action. At the same time, as we might expect in a behavior geared for environmental shaping, the repertoire is not rigid: individual infants vary widely both in how much they babble and in the relative frequency of their babbled sounds (MacNeilage, Hutchinson, & Lasater, 1981).

We should emphasize that segmental phonetic descriptors are simply a convenient, approximate notation of what a child seems to be doing with its articulators—the only descriptors we have, in the absence of cineradiographic or other quantitative data. We should not infer that the child has independent, articulatory control over consonantal and vocalic portions of a syllable. The syllable, formed by rhythmically opening and closing the mouth, is a natural, cohesive unit of speech, with temporal properties that may be determined, in part, by the resonant frequency of the jaw. Its articulatory structure is perhaps related—at least by analogy, if not by homology—to the soft, tongue- or lip-modulated patterns of sound observed in the intimate interactions of Japanese macaque monkeys (Green, 1975; MacNeilage, personal communication).

EARLY PERCEPTUOMOTOR DEVELOPMENT

Imitation, long thought to be the outcome of a lengthy course of cognitive development (e.g., Piaget, 1962), is now known to be an innate capacity of the human infant. Meltzoff and Moore (1977, 1983) have shown, in a pair of meticulously controlled studies, that infants, within 72 hours of birth, can imitate arbitrary facial gestures (mouth opening, lip protrusion) and within 12-21 days (perhaps earlier, but we have no data) can also imitate tongue protrusion and sequential closing of the fingers (of particular interest for sign language acquisition). Of course, these are

relatively crude gestures, far from the subtly interleaved patterns of movement, coordinated across several articulators, that are necessary for adult speech. The importance of the work lies in its implication that optically conveyed, facial gestures, already at birth, induce a neural structure isomorphic with the movements that produce them.

We should not expect speech sounds to induce an analogous neuromotor control structure at birth, not only because the sounds are complex, but because, as language diversity attests, speech is learned. Nonetheless, we might reasonably predict an early, amodal, *perceptual* representation of speech, since this must be the ground on which imitation is based. At present, we have to wait until 4-5 months for this, perhaps because appropriate studies have not yet been done on younger infants. Kuhl and Meltzoff (1982) showed that infants of this age looked longer at the videotaped face of a woman repeatedly articulating the vowel they were hearing (either [i] or [a]) than at the same face articulating the other vowel *in synchrony*. The preference disappeared when the signals were pure tones, matched in amplitude and duration to the vowels, so that infant preference was evidently for a match between a mouth shape and a particular spectral structure. Since spectral structure is directly determined by the resonant cavities of the vocal tract, and since the shape and volume of these cavities are determined by articulation (including pattern of mouth opening for [i] and [a]), the correspondence between mouth shape (optic) and spectral structure (acoustic) reflects their common source in articulation. Evidently, infants of 4-5 months, like adults in recent studies of lip-reading (e.g., McGurk & MacDonald, 1976; Summerfield, 1979, 1987; Campbell, Chapter 7) already have an amodal representation of speech, closely related to the articulatory structures that determine phonetic form.

Just how close this relation is we may judge from a second study similar to that of Kuhl and Meltzoff (1982). MacKain, Studdert-Kennedy, Spieker, and Stern (1983) showed that 5- to 6-month-old infants preferred to look at the videotaped face of a woman repeating the disyllable they were hearing (e.g., [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g., [vava]). However, the two faces were presented to left and right of an infant's central gaze, and the preference for an acoustic-optic match was only significant when infants were looking at the right side display. We may interpret this result in light of studies by Kinsbourne and his colleagues (e.g., Kinsbourne, 1972; Lempert & Kinsbourne, 1982), demonstrating that attention to one side of the body facilitates processes for which the contralateral hemisphere is specialized. Infants might then be more sensitive to acoustic-optic correspondences in speech presented on their right sides than on their left. Thus, infants of five to six months may already have an amodal representation of speech in the hemisphere that will later coordinate the activity of their bilaterally innervated speech apparatus.

Signally absent from all of the foregoing is any indication that the infant is affected by the surrounding language. In fact, it has often been proposed (e.g., Brown, 1958) that the infant's phonetic repertoire drifts towards that of its native language during the babbling of the second half year, but, despite several studies, no firm evidence of babbling drift has been found (Locke, 1983). We do, however, have evidence of perceptual effects. Werker and her colleagues (Werker, 1982; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984) have shown, in several cross-sectional and longitudinal studies, that, during the second half year, infants may gradually lose their capacity to distinguish sound contrasts not used in their native language. This is perhaps just the period when an infant is first attending to individual words and the situations in which they occur (Jusczyk, 1982; MacKain, 1982).

The general picture of perceptuomotor development over the first year, then, is of two parallel, independent processes, with production trailing perception. Doubtless, physiological changes in the left hemisphere are taking place, laying down neural networks that will later make contact. These processes may resemble those in songbirds, such as the marsh wren, in which the perceptual template of its species' song is laid down during a narrow sensitive period many weeks before it begins to sing (Kroodsma, 1981). The first behavioral evidence of a perceptuomotor link then appears with the bird's first song and, in the infant, with its first imitation of an adult sound.

FIRST WORDS AND THEIR COMPONENT GESTURES

Up to this point we have talked easily of perceptual, or perceptuomotor, "representations" without asking what is represented. During the 1970s, when intensive work on child phonology began, researchers quite reasonably assumed that units of acquisition would be those that linguists had found useful in describing adult language: features and phonemes. Little attention was paid to the fact that these units, as defined by linguists, were abstract descriptors that could not be specified either articulatorily or acoustically, and were therefore of dubious utility to the child striving to talk like its companions. The oversight was perhaps encouraged by division of labor between students of perception and students of production whose mutual isolation absolved them from confronting what the child confronts: the puzzle of the relation between listening and speaking.

Over the past decade, child phonologists have come to recognize the fact, borne in also by pragmatic studies (e.g., Bates, 1979), that a child's entry into language is mediated by meaning; and meaning cannot be conveyed by isolated features or phonemes. The child's earliest unit of meaning is probably the prosodic contour: the rising pitch of question and surprise, the falling pitch of declaration, and so on, often observed in stretches of "jargon" or intonated babble (Menn, 1978). The earliest *segmental* unit of meaning is the word (or formulaic phrase).

Evidence for the word as the basic unit of contrast in early language is rich and subtle (Ferguson, 1978; Ferguson & Farwell, 1975; Macken, 1979; Menn, 1983a; Moskowitz, 1973). Here I simply note three points. First is the observation that phonetic forms mastered in one word are not necessarily mastered in another. For example, a 15-month-old child may execute [n] correctly in *no*, but substitute [m] for [n] in *night*, and [b] for [m] in *moo* (Ferguson & Farwell, 1975). Thus, the child does not contrast [b], [m] and [n], as in the adult language, but the three words with their insecurely grasped onsets.

A second point is that early speech is replete with instances of consonant harmony, that is, words in which one consonant assimilates the place or manner of articulation of another—even though the child may execute the assimilated consonant correctly in other words. Thus, a child may produce *daddy* and *egg* correctly, but offer [gɔg] for dog and [dʌt] for *duck* and *truck;* the child seems unable to switch place of articulation within a syllable. Such "assimilation at a distance" suggests that the word is "assembled before it is spoken" as a single prosodic unit (Menn, 1983a, p. 16).

The third point is that individual words may vary widely in their phonetic form from one occasion to another. A striking example comes from Ferguson and Farwell (1975). They report ten radically different attempts by a 15-month-old girl, K, to say *pen* within one half-hour session:

[mə̃^{ə}, ^{v}ʌ̃, de^{dn}, hin, ^{m}bõ, p^{h}in, t^{h}n̩t^{h}n̩t^{h}n̩, ba^{h}, d^{h}au^{n}, buə̃].[2]

On the surface, these attempts seem almost incomprehensibly diverse one from another and from their model. But the authors shrewdly remark that "K seems to be trying to sort out the features of nasality, bilabial closure, alveolar closure, and voicelessness" (Ferguson & Farwell, 1975, p. 14). An alternative description (to be preferred, in my view, for reasons that will appear shortly) would be to say that all the *gestures* of the model (lip closure, tongue raising and fronting, alveolar closure, velum lowering/raising, glottal opening/closing) are to be found in one or other of these utterances, but that the gestures are incorrectly phased with respect to one another. For example, lip closure for the initial [p], properly executed with an open glottis and raised velum, will yield [^{m}b], as in [^{m}bõ], if glottal closure for [ɛn] and velum lowering for [n] are initiated at the same time as lip closure, tens of milliseconds earlier than in the correct utterance [pɛn]. Thus, the adult model evidently specified for the child the required gestures, but not their relative timing. (Notice, incidentally, that the only gestures present in the child's attempts, but absent from the model, are tongue backing and tongue lowering for the sounds transcribed as [o], [a], and [u]. Four of these five "errors" occur when the child has successfully executed initial lip closure, as though attention to the initial gesture had exhausted the child's capacity to assemble later gestures.)

One reason for preferring a gestural to a featural description of a child's—or for that matter of an adult's—speech is that it lends the description observable, physical content (Browman & Goldstein, 1986). We are then dealing with patterns of movement in space and time, accessible to treatment according to general principles of motor control (e.g., Kelso, Tuller, & Harris, 1983; Saltzman & Kelso, 1987). For example, the problem of motor equivalence may become more tractable, because the gesture is a *functional* unit, an equivalence class of coordinated movements that achieve some end (closing the lips, raising the tongue, etc.) (Kelso, Saltzman, & Tuller, 1986). Moreover, a gestural description may help us to explore the claim (based on the facts of imitation) that the speech percept is an amodal structure isomorphic with the speaker's articulation. Glottal, velic, and labial gestures can already be isolated by standard techniques; tongue movements are more problematic, because they are often vectors of two or more concurrent gestures.

Nonetheless, positing a concrete, observable event as the fundamental unit of production may help researchers to analyze articulatory vectors into their component forces, and to isolate the acoustic marks of those vectors in the signal.

Finally, to forestall misunderstanding, a gestural description is not simply a change of terminology. Gestures do not usually correspond one-to-one with either phonemes or features. The phoneme /m/, for example, comprises the precisely timed and coordinated gestures of bilabial closure, velum lowering and glottal closing. The gesture of bilabial closure corresponds to several features [- continuant], [+ anterior], [+ consonantal], etc. A gestural account of speech—that is, an account grounded in the anatomy and physiology of the speaker—will require extensive revision of standard featural or segmental descriptions (Browman & Goldstein, 1986).

FROM WORDS TO PHONEMES

To summarize the previous section, we have argued that: (1) an element of meaning, the word, is the initial segmental unit of contrast in early speech; (2) a word is a coordinated pattern of gestures; (3) an adult spoken word specifies for the child learning to speak, at least some of its component gestures, but often not their detailed temporal organization. (The third point does not imply that the child's perceptual representation is necessarily incomplete: the representation may be exact, and the child's difficulty solely in coordinating its articulators. The difference is not without

theoretical interest, but, in the present context, our focus is on how a child comes to reorganize a holistic pattern of gestures into a sequence of phonetic segments, or phonemes. Whether the reorganization is perceptual, articulatory, or both need not concern us).

What follows, then, is a sketch of the process by which phonemes seem to emerge as units of perceptuomotor control in a child's speech. I should emphasize that details of the process vary widely from child to child, but the general outline is becoming clear (Menn, 1983a, 1983b).

We can illustrate the process by tracing how a child escapes from consonant harmony, that is, how it comes to execute a word (or syllable) with two different places (or manners) of articulation. Children vary in their initial attack on such words: Some children omit, others harmonize one or other of the discrepant consonants. For example, faced with the word *fish*, which calls for a shift from a labiodental to a palatal constriction, one child may offer [fi'], another [ij]; faced with *duck*, one child may try [gʌk], another [dʌt]. Menn (1983b) proposes a perspicuous account of such attempts: the child has "...learned an articulatory program of opening and closing her mouth that allows her to specify two things: the vowel and one point of oral closure" (p. 5). Reframing this in terms of gestures (an exercise that we need not repeat in later examples), we may say that the child has learned to coordinate glottal closing/opening and tongue positioning (back, front, up, down, in various degrees) with raising/lowering the jaw, in order to approximate an adult word. This description of a word as an articulatory program, or routine, composed of a few variable gestures, is a key to the child's phonological development.

Consider here a Spanish child, Si, studied by Macken (1979) from 1 year 7 months to 2 years 1 month of age. At a certain point, Si seemed only able to escape from consonant place harmony by producing a labial-vowel-dental-vowel disyllable, deleting any extra syllable in the adult model. Thus, *manzana* ('apple') became [manna]; *Fernando* became [mannə] or [wanno], with the initial [f] transduced as [m] or [w]; *pelota* ('ball') became [patda]. In some words, where the labial and dental were in the "wrong" order, Si metathesized. Thus, *sopa* ('soup') became [pwæta], replacing [s] with [t], and *teléfono* became [fəntonno]. As Si's mastery increased, the class of words, subject to the labial-dental routine, narrowed: *manzana* became [tʃænna], *Fernando* became [tçınalto], and so on.

These examples make two important points: (1) the child brings adult words with similar patterns (e.g., *manzana, Fernando, pelota)* within the domain of a single articulatory routine, demonstrating use of the word as a unit; (2) at the same time, the child selects as models adult words that share certain gestural patterns, demonstrating an incipient grasp of their segmental structure.

We may view the developmental process as driven by the conflicting demands of articulatory "ease" and lexical accumulation. As long as the child has only a few words, it needs only one or two articulatory routines. Initially, it exploits these routines by adding to its repertoire only words composed of gestural patterns similar to those it has already "solved," and by avoiding words with markedly different patterns. (For evidence and discussion of avoidance and exploitation in early child phonology, see Menn, 1983a.) Once the initial routines have been consolidated, new routines begin to emerge under pressure from the child's accumulating vocabulary. New routines emerge either to handle a new class of adult words, not previously attempted, or to break up and redistribute the increasing cohort of words covered by an old articulatory routine.

Phonological development seems then to be a process of: (1) diversifying articulatory routines to encompass more and more different classes of adult model; (2) gradually narrowing the domain within a word to which an articulatory routine applies. The logical end of the process (usually reached during the third year of life,

when the child has accumulated some 50-100 words) is a single articulatory routine for each phonetic segment. Development is far from complete at this point: there must ensue, at least, the systematic grouping of phonetic variants (allophones) into phoneme classes, and the discovery of language-specific regularities in their sequencing ("phonotactic rules"). But the emergence of the phonetic segment as a perceptuomotor unit brings the entire adult lexicon, insofar as it is cognitively available, within the child's phonetic reach. This signals the onset of the explosive vocabulary growth, at an average rate of some 5-7 words a day, that permits an average 6-year-old American child to recognize an estimated 7,000-11,000 root words, depending on family background (Templin, 1957; cf. Miller, 1977).

CONCLUSION

We began with a paradox; the apparent incommensurability of the quasi-continuous articulatory and acoustic structure of speech with the discrete units of its written representation. To resolve the paradox, we proposed that an alphabetic letter (or an element in a syllabary, or an ideograph) is a symbol for a discrete, perceptuomotor control structure. We then traced the emergence of such structures as encapsulated patterns of gesture in a child's speech. Implicit in their derivation is that a child, once possessed of them, is, at least in principle, ready for the metalinguistic task of learning to write and read (cf. Asbell, 1984).

What we have left unresolved is the relation between discrete motor control structures (word, syllable, phoneme) and the coordinated patterns of gesture that they generate. Perhaps we should regard the postulated structures as conceptual place-holders. Their functional analysis must await advances in neurophysiology and in the general theory of motor control.

ACKNOWLEDGMENT

My thanks to Björn Lindblom and Peter MacNeilage for critical comments and discussion. The paper was written while the author was on sabbatical leave as a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. The financial support of the City University of New York and of the Spencer Foundation is gratefully acknowledged. Some of the research reported received support from NICHD grant HD 01994 to Haskins Laboratories.

REFERENCES

Alegria, J., & Noirot, E. (1982). Oriented mouthing activity in neonates: Early development of differences related to feeding experiences. In J. Mehler, S. Franck, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities* (pp. 389-397). Hillsdale, NJ: Erlbaum.

Asbell, B. (1984). *Writer's workshop at age 5*. New York Times Magazine, February 26th.

Bates, E. (1979). *The emergence of symbols*. New York: Academic.

Bellugi, U., Poizner, H., & Klima, E. S. (1983). Brain organization for language: Clues from sign aphasia. *Human Neurobiology, 2*, 155-170.

Best, C. T., Hoffman, H., & Glanville, B. R. (1982). Development of infant ear asymmetries for speech and music. *Perception & Psychophysics, 31*, &5-85.

Blumstein, S. E. (1981). Phonological aspects of aphasia. In M. T. Sarno (Ed.), *Acquired aphasia*. New York: Academic.

Browman, C. P. (1980). Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Brown, R. (1958). *Words and things*. Glenco, IL: Free Press.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum.

Cowan, N., Leavitt, L. A., Massaro, D. W., & Kent, R. D. (1982). A fluent backward talker. *Journal of Speech and Hearing Research, 25*, 48-53.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mother's voices. *Science, 208*, 1174-1176.

Eimas, P. D. (1985). The perception of speech in early infancy. *Scientific American, 252*, 46-52.

Fentress, J. C. (1981). Order in ontogeny: Relational dynamics. In K. Immelmann, G. W. Barlow, L. Petrinovich, & M. Main (Ed.), *Behavioral development* (pp. 338-371). New York: Cambridge University Press.

Fentress, J. C. (1983). Hierarchical motor control. In M. Studdert-Kennedy (Ed.), *Psychobiology of language* (pp. 40-41). Cambridge, MA: MIT Press.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language, 51*, 419-430.

Ferguson, C. A. (1978). Learning to pronounce: The earliest stages of phonological development in the child. In F. D. Minifie & L. L. Lloyd (Eds), *Communicative and cognitive abilities—Early behavioral assessment* (pp. 273-297). Baltimore, MD: University Park Press.

Golani, I. (1981). The search for invariants in motor behavior. In K. Immelmann, G. W. Barlow, L. Petrinovich, & M. Main (Eds), *Behavioral development* (pp. 372-390). New York: Cambridge University Press.

Green, S. (1975). Variation of vocal pattern with social situation in the Japanese monkey (Macaca fuscata): A field study. In L. A. Rosenblum (Ed.), *Primate behavior* (Vol. 4, pp. 1-102). New York: Academic.

Jusczyk, P. W. (1982). Auditory versus phonetic coding of speech signals during infancy. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), *Perspectives on mental representation* (pp. 361-387). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14*, 29-59.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech* (pp. 138-173). New York: Springer-Verlag.

Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. *Science, 176*, 539-541.

Kroodsma, D. E. (1981). Ontogeny of bird song. In G. B. Barlow, K. Immelmann, M. Main, & L. Petrinovich (Eds.), *Behavioral development* (pp. 518-532). New York: Cambridge University Press.

Kuhl, P. K. (1986). Infants' perception of speech: Constraints on the characterizations of the initial state. In B. Lindblom & R. Zetterström (Eds), *Precursors of early speech* (pp. 219-244). Basingstoke, UK: MacMillan.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138-1144.

Lempert, H., & Kinsbourne, M. (1982). Effect of laterality of orientation on verbal memory. *Neuropsychologia, 20*, 211-214.

Levinson, S. E., & Liberman, M. Y. (1981). Speech recognition by computer. *Scientific American, 244*, 64-76.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461

Lindblom, B., MacNeilage, P. F., & Studdert-Kennedy, M. (forthcoming). *Evolution of spoken language*. Orlando, FL: Academic.

Locke, J. (1983). *Phonological acquisition and change*. New York: Academic.

MacKain, K. S. (1982). Assessing the role of experience in infant speech discrimination. *Journal of Child Language, 9*, 527-542.

MacKain, K. S. (1983). Speaking without a tongue. *Journal of the National Student Speech Language Hearing Association, 11*, 46-71.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science, 219*, 1347-1349.

MacKay, D. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323-350.

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua, 49*, 11-49.

MacNeilage, P. F., Hutchinson, J., & Lasater, S. (1981). The production of speech: Development and dissolution of motoric and premotoric processes. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 503-519). Hillsdale, NJ: LEA.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 175-178.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development, 54*, 702-709.

Menn, L. (1978). *Pattern, control, and contrast in beginning speech: A case study in the development of word form and function*. Bloomington, IN: Indiana University Linguistics Club.

Menn, L. (1983a). *Development of articulatory, phonetic, and phonological capabilities*. In B. Butterworth (Ed.), Language production, Vol. II. London: Academic.

Menn, L. (1983b). *Language acquisition, aphasia and phonotactic universals*. Paper presented at 12th Annual University of Wisconsin-Milwaukee Linguistics Symposium.

Miller, G. A. (1977). *Spontaneous apprentices*. New York: The Seabury Press.

Molfese, D. L. (1977). Infant cerebral asymmetry. In S. J. Segalowitz & F. A. Gruber (Eds.), *Language development and neurological theory* (Vol. 4, pp. 29-59). New York: Academic Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition, 7*, 323-331.

Moskowitz, A. I. (1973). The acquisition of phonology and syntax. In K. K. J. Hintikka, J. M. E. Moravsik, & P. Suppes (Eds.), *Approaches to natural language* (pp. 48-84). Boston: D. Reidel.

Neville, H. J. (1980). Event-related potentials in neuropsychological studies of language. *Brain and Language, 11*, 300-318.

Neville, H. J. (1985). Effects of early sensory and language experience on the development of the human brain. In J. Mehler & R. Fox (Eds.), *Neonate cognition* (pp. 349-363). Hillsdale, NJ: Erlbaum.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology*. (Vol. 1: Production, pp. 93-112). New York: Academic Press.

Oller, D. K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 21-35). Basingstoke, UK: MacMillan.

Piaget, J. (1962). *Play, dreams, and imitations in childhood*. New York: W. W. Norton.

Pisoni, D. B. (1985). Speech perception: Some new directions in research and theory. *Journal of the Acoustical Society of America, 78*, 381-388.

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. *Psychological Review, 94*, 84-106.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 58-76). Hillsdale, NJ: Erlbaum.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*, 314-331.

Summerfield, Q. (1987). Preliminaries to a comprehensive account of audiovisual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye* (pp. 3-51). Hillsdale, NJ: Erlbaum.

Templin, M. C. (1957). *Certain language skills of children*. Minneapolis: University of Minnesota Press.

Turvey, M. T., & Kugler, P. N. (1984). A comment on equating information with symbol strings. *American Journal of Physiology, 246*, 925-927.

Welty, E. (1983). *One writer's beginnings*. New York: Warner Books.

Werker, J. F. (1982). *The development of cross-language speech perception: The effect of age, experience and context on perceptual organization*. Unpublished doctoral dissertation, University of British Columbia.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

FOOTNOTES

[1]In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67-84). London: Academic Press, 1987.

[†]Also the University of Connecticut and Yale University.

[1]Speech errors display a number of well-known biases. For example, word-initial errors are more common than word-medial and word-final errors; metathesis occurs only between segments that occupy the same position in the syllabic frame; and the phonetic form of an error is adjusted to the context in which the erroneous segment occurs, not to the context from which it was drawn. Thus, speech errors often reflect phonetic processes that follow access of a phonemically specified lexical item. An adequate account of speech errors must therefore accommodate not only the phoneme as the fundamental phonological unit of all spoken languages, but the processes of lexical access and phonetic execution that give rise to biases in speech error types and frequencies. For a model of speech errors within these constraints, see Shattuck-Hufnagel (1983); for fuller discussion of the issue, see Lindblom, MacNeilage, and Studdert-Kennedy (forthcoming).

[2]The first two items listed were immediate imitations of an adult utterance. Later items were identified by their "...consistency in reference or accompanying action" (Ferguson & Farwell, 1975, p. 9). Interobserver agreement in the study from which these transcriptions were drawn was over 90%. The validity of the assumed target, *pen*, is further attested by the many featural (or gestural) properties common to the target and each of the child's attempts. The attempts did not include, for example [gog], in which only glottal closure would be shared with the presumed target.

Consonant-vowel Cohesiveness in Speech Production as Revealed by Initial and Final Consonant Exchanges*

Carol A. Fowler†

Two experiments use a procedure developed by Carter and Bradshaw (1984) to examine the role of syllable structure in speech production. In the procedure, subjects exchange phonological segments in corresponding positions of a pair of visually-presented words or nonwords and to produce the resulting words or nonwords as quickly as possible. Carter and Bradshaw had shown that the pattern of latencies mirror that of frequencies of exchange errors in natural speech. The first experiment of the present study shows that initial consonant exchanges are promoted by phonetic similarity of the exchanging consonants and they reflect a bias for producing real words. With these influences controlled, Experiment 2 replicates and extends the finding of Carter and Bradshaw that initial consonant exchanges are made more rapidly than final consonant exchanges. The discussion relates the latency difference between these conditions to a difference in the "cohesiveness" of initial and final consonants with their vowel. In particular, in Experiment 2, more than one-third of errors made on final-consonant or vowel exchanges are exchanges of the whole syllable rhyme (VC), whereas just 10% of errors made on initial consonant or vowel exchanges are exchanges of the initial CV of the word. Various explanations for the difference in cohesiveness are examined in post hoc analyses.

INTRODUCTION

Spontaneous errors of speech production (for example, "morage in the fountains" for "forage in the mountains" or "even the best team losts" for "even the best teams lost") have proven quite revealing sources of information concerning the structures and processes involved in language production (see, for example, Dell, 1986; Garrett, 1976, 1980; Shattuck-Hufnagel, 1979, 1983). The errors are strikingly systematic in the units of speech that participate in errors (for example, phonemes and words, but rarely syllables and features) and in the conditions that promote errors. Recently, however, researchers have recognized that converging evidence is needed from laboratory studies of speech (e.g., Baars & Motley, 1976; Cutler, 1981; MacKay, 1971) because speech error collections are subject to bias and because they do not provide all the information needed to evaluate error patterns. One source of bias in corpora of spontaneous errors is contamination by mishearings. This is less likely to be characteristic of speech transcribed from recordings made in a laboratory setting. As for missing information in error collections, questions relating to the frequency with which certain error types occur in comparison with their opportunities to occur cannot easily be answered using spontaneous-error collections because opportunities for error are not available except by estimate from other sources.

A variety of experimental techniques has been developed (e.g., Baars & Motley, 1971; Carter & Bradshaw, 1984; Dell, 1986; Kupin, 1979; Levitt & Healy, 1985; MacKay, 1971) to study language production and speech errors in the laboratory. These procedures have verified many of the characteristics of errors previously observed in corpora of spontaneous speech errors. But in addition, they have uncovered new ones that are difficult to notice in spontaneous-error corpora.

For example, many researchers now have identified a "lexical bias" in experimentally-elicited phoneme errors such that errors are more likely to occur if the erroneously produced string is a real word of the language than if they are not (Baars, Motley, & MacKay, 1975; Dell, 1980). In contrast, based on the high proportion of nonwords that they had observed in their error corpora, Fromkin (1973) and Garrett (1975, 1976) had concluded that phoneme errors do not tend to create real words. Uncovering a statistical tendency such as a lexical bias is made easier if experimental conditions can be established in which opportunities for word- and non-word-creating phoneme errors are controlled. (But see Dell [1980; Dell & Reich, 1981], who demonstrates that the same bias does occur in spontaneous phonological-segment errors as well.)

Another area in which laboratory studies of speech errors have made a special contribution is in our understanding of the conditions that affect phoneme-substitution patterns. Examination of substituting and substituted phonemes in an error corpus had led Shattuck-Hufnagel and Klatt (1979) to conclude that, in general, errors are symmetrical such that the frequency with which, for example, /z/ substitutes for /d/ is nearly the same as the frequency with which /d/ substitutes for /z/. In turn, the symmetry suggested that substituting phonemes are not, in any sense (such as their relative frequency in the language or their relative markedness), "stronger" than substituted phonemes. However, as Levitt and Healy (1985) point out, this conclusion involves the implicit assumption that the opportunities for /d/ and /z/ to be targets of substitutions is the same. The opportunities are the same in cases of errors in which both the substituting and substituted segments are present in the intended utterance, but they are not when the substituting segment is not in the utterance as planned. Instead, /d/ occurs more frequently than /z/ in speech (Roberts, 1965). Thus, in order for the *number* of /z/ → /d/ and /d/ → /z/ substitutions to be the same, the *probability* of /d/ substituting for /z/ must in fact be higher than the probability of /z/ substituting for /d/. In experiments controlling opportunities for error, Levitt and Healy were able to support hypotheses that some segments are stronger than others in that they participate as substituting phonemes relatively frequently and that strength is in part a function of the segments' frequencies of usage in the language. Baars and Motley (1975) show, in addition, that strength is related to markedness and to the transition frequencies of phoneme sequences created by an error.

The present research adopts an experimental approach to the study of the role of syllable structure in speech errors and in speech production more generally (see also MacKay, 1978). It follows a recent study by Carter and Bradshaw (1984).

In spontaneous errors, pairs of segments involved in errors (e.g., "morage in the fountains") almost always preserve their intended position in the syllable. That is, with few exceptions syllable-initial consonants involved in movement errors exchange with other syllable-initial consonants; vowels exchange only with vowels, and final consonants with final consonants. Moreover, by far the most frequent single-segment errors involve initial consonants of syllables and words; final consonant errors are reported with a much lower frequency. Spontaneous-error corpora allow at least three interpretations of this imbalance in error frequencies. In part it may be due to differential likelihoods that error collectors will hear initial- as compared to final-consonant errors. Listeners are known to "restore" incorrectly-

produced phonemes later in a word with greater frequency than they restore initial phonemes (Marslen-Wilson & Welsh, 1978). A second possibility is that whereas errors in initial and final position of a syllable or word may occur under exactly the same conditions (e.g., the interacting consonants tend to be phonetically similar, occur in similar environments [Shattuck-Hufnagel, 1979], and create real words of the language), other things equal, those conditions may arise with greater frequency in syllable-initial than in syllable-final positions. A third possibility is that, as other theorists and researchers have suggested for independent reasons, syllables may have an internal structure in which vowels and final consonants cluster into a constituent (called the "rime" or "rhyme"), while initial consonants serve as their own constituent, the "onset." One reflection of this constituent structure may be a greater cohesion of final consonants with their vowels than of initial consonants and vowels and therefore, a lesser tendency for final consonants to break off from vowels in errors. Other reflections are that poetic devices (such as alliteration on the one hand and rhyming on the other) respect the organization of syllables into onsets and rhymes, popular language games (such as Pig Latin) do, stress rules of English and many languages refer to syllable rhymes (see e.g., Prince, 1980) but not to onsets, and both children (Treiman, 1985) and adults (Treiman, 1986) learn word games more easily that involve segmenting or blending syllables along constituent-structure lines than elsewhere.

Carter and Bradshaw (1984) have developed an experimental procedure to exploit the possibility that the relative frequency with which segments participate in errors relates to their degree of detachability from a syllable or word. They explicitly asked subjects to exchange two phonemic segments in corresponding positions in a word pair and measured the latency and accuracy with which subjects were able to do so. The segments to be exchanged were, in different conditions, the initial consonants, the medial vowels and the final consonants of pairs of monosyllables presented visually. If both the frequency with which segments participate in spontaneous errors and the latency with which they can be exchanged on purpose redundantly reflect the detachability of a segment from a syllable or word, then latencies should pattern as errors do. In support of this idea, Carter and Bradshaw found two latency patterns characteristic of speech error patterns. First, they found a lexical bias in intended segment exchanges such that exchanges creating real words were made more rapidly than those creating nonwords. Second, they found that initial consonants were exchanged more rapidly than medial vowels or final consonants. Because the dominance of initial consonant exchanges appeared in the latency measure, the possibility suggested earlier that initial consonant errors predominate in spontaneous errors only because they are more likely to be detected by collectors can be ruled out.

The present experiments used a modification of Carter and Bradshaw's procedure to pursue the findings on syllable structure. The experiments of Carter and Bradshaw leave open the possibility also suggested earlier that, other things equal, contextual conditions may tend to favor initial over final consonant exchanges. Although Carter and Bradshaw controlled the number of opportunities for initial- and final-consonant errors, they did not explicitly control for other variables that may generally differ between consonants in those positions in a syllable or word. In particular, consonant exchanges in spontaneous errors appear to be promoted by phonetic (featural) similarity between the exchanging phonemes and by phonetic similarities in the environments of the exchanging segments (Shattuck-Hufnagel, 1979). In Experiment 1 below, using a slight modification of the paradigm of Carter and Bradshaw, I confirm that phonetic similarity of exchanging consonants promotes exchanges. In Experiment 2, I control the similarity of initial and final

consonants and the environments in which they occur and ask whether their exchange latencies differ.

If differential phonetic similarity does not explain the tendency for initial consonants to dominate in single phoneme exchanges, a different possibility is that the tendency indeed reflects an onset-rhyme constituent structure of the syllable. This interpretation presumably is favored by Carter and Bradshaw and promoted also by the independent evidence on syllable structure cited earlier. Direct evidence for this interpretation might be provided by the pattern of errors in the exchange-latency paradigm. If final consonants are, in some sense, more cohesive with the vowel than are initial consonants, then on occasions in which subjects are instructed to exchange vowels, for example, mistakes in which they exchange the whole rhyme should be relatively more frequent than mistakes in which they exchange the initial consonant-vowel (CV) sequence. Likewise, instructed to exchange initial consonants, subjects should inadvertently exchange CVs rarely; however, instructed to exchange final consonants, they should exchange VCs relatively frequently. In Experiment 2, I look for such error patterns in subjects' exchange errors.

An explanation of speech errors and latency in terms of syllable structure or relative "cohesiveness" is, of course, incomplete without some indication of what work, exactly, syllable structure does in language production. Dell (1986) has proposed, for example, that the constituent structure of syllables is reflected explicitly in associative connections between phonemic segments and abstract onset and rhyme nodes in a spreading-activation network representing the lexicon. These explicit connections foster cohesiveness among the subunits sharing a common higher node. However, it would be useful to discover *why* syllables structure in that way, particularly in view of suggestions that many languages, indeed the majority that have been examined (Prince, 1980), have an onset-rhyme structure. Two post-hoc analyses of the findings of Experiment 2 are performed to examine possible sources or correlates of the effects of syllable structure on speech errors.

EXPERIMENT 1

The first experiment was conducted in part to look for effects of phonetic similarity of consonant segments on time to make an initial-consonant exchange and in part to test a small modification to the procedure of Carter and Bradshaw. Behind these aims of the experiment was an attempt, in addition to those of Carter and Bradshaw, to verify that the exchange-latency paradigm yields data that pattern similarly to frequency patterns of natural speech errors.

In the research by Carter and Bradshaw, subjects were timed as they performed segment exchanges in a list of letter-string pairs. The dependent measure was the average time per stimulus pair to make an exchange, obtained by dividing the time to make all of the exchanges in the list by the number of pairs in the list. Because the procedure required subjects to read a pair of items, cover it with a card, and then report the exchange, latencies included the times to execute all of these activities. Latencies were slow and presumably more variable than they might be if some of these components were excluded from the latency measure. In an effort to remove the interval to utter the response pair from the latency measure (that is, to remove time from articulation onset to offset), Carter and Bradshaw measured articulation time separately and subtracted it from the latency measure. One consequence of this correction was to eliminate the previously significant effects of response-pair lexicality on response time. If this procedure indeed successfully removes articulation time from the latency measure and removes nothing else, then it shows that response pair lexicality affects only articulation time, not time to exchange

consonants or vowels. However, it would also indicate a finding using the exchange-latency procedure that differs from findings on speech errors. As already noted, in both error corpora and experimentally-elicited errors, there is a lexical bias in the frequency of phoneme exchanges.

The modification to the procedure adopted here was to obtain a measure from which articulation time was excluded from the outset. The procedure was to obtain a measure of time from visual presentation of a single stimulus pair to the onset of the acoustic signal for the subject's exchange response. This was accomplished by presenting stimulus pairs on a computer-terminal screen and obtaining response times using a voice key. Another benefit of this modification may be to reduce some of the variability of the average latency measure of Carter and Bradshaw due to the subject's variable times to cover each stimulus pair after reading it.

Using this modified procedure, then, I asked whether effects of lexicality of the stimulus and response strings could be detected in the latency measure and whether phonetic similarity of the exchanged consonants reduces latency to exchange the pair.

Method

Subjects. Subjects were 18 students at Dartmouth College who participated in the experiment for course credit. They were native speakers of English who reported normal speech and hearing.

Materials. Materials consisted of 72 pairs of letter strings, 12 each in the six cells representing the crossing of three levels of lexicality of stimulus and response strings described below and two levels of phonetic similarity of the consonants involved in the exchange. Stimulus pairs are listed in Appendix 1. In one lexicality condition, W-W, both the stimulus and response strings were real words of English (e.g., PASTE TOLL → TASTE POLL); in a second condition, W-NW, the stimulus strings were words, but response strings were pseudowords (e.g., PAIN TOAD → TAIN POAD); in the third condition, stimulus strings were pseudowords and response strings were words (e.g., PAME TOPE → TAME POPE). Across conditions having real-word stimulus strings, words were matched as closely as possible in median frequency according to the tables of Francis and Kučera (1982). (The range of median frequencies across the six conditions was 12-26 occurrences in the Brown corpus.) Likewise, across conditions having real-word response strings, response words were matched in median frequency (range: 8-16 occurrences).

Twelve pairs of consonants were selected to serve as initial consonants in the condition in which consonants of a stimulus pair were phonetically similar. The same 12 consonant pairs were used at all three levels of lexicality. The consonants differed by one phonetic feature according to the feature system of van den Broecke and Goldstein (1980). This feature system was used because Levitt and Healy (1985) had found that in comparison with the feature system of Chomsky and Halle (1968), it accounted well for phoneme-error frequency patterns.

To create the condition in which consonants were phonetically dissimilar, I re-paired the consonants in the phonetically-similar condition so that new pairs differed by at least two phonetic features. There was just one exception to this in which a /t/ in the similar condition was changed to /s/ in the dissimilar condition to create a two-feature difference in a new stimulus pair of initial consonants.

An attempt was also made to use the same medial vowels across the six experimental conditions, or, in three instances, where that was not possible, to use vowels with about the same degree of opening. This precaution and the fact that the set of consonants triggering the voice key was matched across conditions obviated any need to measure and correct for voice key effects. A set of 24 practice trials was

also devised, consisting of words and pseudowords similar to those used in the test trials.

Procedure. Subjects were run individually. They were instructed that they would see pairs of letter strings presented on a computer-terminal screen. They were to exchange the initial consonant sounds of the pair members as quickly and as accurately as possible and to say the response string into the microphone in front of them. The distinction between exchanging initial *sounds* of words rather than *spellings* was illustrated by example, using words having initial consonants spelled with two letters (such as "th") and using words having initial consonant letters (e.g., "c") whose pronunciation might change were subjects to exchange letters rather than sounds.

Subjects were instructed not to begin to say the exchanged pair until they had determined the pronunciation of both members of the pair. To encourage them to do this (and, therefore, to obtain latencies that consistently reflected time to make the whole exchange), the stimulus string was line fed off the screen as soon as the voice-key was triggered by the subject's initial response.

Letter strings were printed in capital letters on the terminal screen out of sight of the subject and then were line fed into view as the response-time clock began to measure latency to respond. The experimenter sat opposite the subject, viewing a monitor that showed both the stimulus items and the correct response items. During the block of 24 practice trials, the experimenter provided corrections to the subject as needed. Following the 24 practice trials, subjects were given feedback on the terminal screen consisting of their average response time for the block. Next they received three blocks of 24 test trials with response-time feedback after each block, but with no experimenter-provided corrective feedback. Experimental conditions were randomized within blocks and were differently randomized for each subject.

Subjects' responses were recorded on audio tape for later checking. Failures of the voice key to trigger were eliminated from the data. In addition, erroneous responses were eliminated from analysis of latencies.

Design. The experiment had two independent variables, lexicality of stimulus and response strings, with three levels, W-W, NW-W and W-NW, and phonetic similarity of initial consonants, with two levels. Dependent measures were latency and percent errors.

Results and Discussion

Latencies and error percentages for the three lexicality conditions and the two levels of phonetic similarity are given in Table 1. Errors averaged just over 10% in all conditions and showed no significant differences across conditions in an analysis of variance. They will not be considered further.

In an analysis of variance with factors, lexicality and phonetic similarity, both independent variables were significant: lexicality, $F(2,34) = 7.64$, $p = .002$; phonetic similarity, $F(1,34) = 4.18$, $p = .05$. The interaction between the variables was not significant, $F(2,34) = 1.70$, $p = .2$. Post hoc analyses showed that the effect of lexicality was due to a difference between the two conditions in which the response items were real words on the one hand and the condition in which they were pseudowords on the other. Conditions in which response items were real words of English led to faster exchanges than conditions in which response items were nonwords. The 68 ms difference between the W-W and NW-W conditions was not significant. Therefore, lexicality of the response did affect latency significantly, while lexicality of the stimulus items did not.[1]

In large part, the effect of lexicality on latency is consistent with findings of Carter and Bradshaw before they attempted to eliminate articulation time from their response-time measure. In addition, it is consistent with the lexical bias

characteristic of error corpora (Dell, 1980) and of experimentally elicited errors (Baars et al., 1975).

TABLE 1

Response Times and Error Percentages for the Experimental Conditions of Experiment 1.

	LEXICALITY		
	W-W	NW-W	W-NW
RT	1561	1629	1802
Error	10.6	10.7	10.3

	PHONOLOGICAL SIMILARITY	
	Similar Pairs	Dissimilar Pairs
RT	1628	1701
Error	10.4	10.7

Phonetic similarity shortened exchange latency as predicted. This finding is consistent with observations from spontaneous error corpora that exchanges occur disproportionately frequently among phonetically similar consonants. Although the effect was reliable, it was weak. Accordingly, it is unlikely to have been a major source of confounding in the research by Carter and Bradshaw even though it was not explicitly controlled in that research.

The present experiment suggests, following the work of Carter and Bradshaw, that the exchange-latency paradigm does lead to latencies that pattern in ways similar to the patterning of relative frequencies of errors on phoneme-error corpora. In the present experiment, this generalization holds true in respect to effects of lexicality of the response strings and the phonetic similarity of segments that interact in exchanges. In Experiment 2, the paradigm is used to examine effects of syllable structure on response latency and errors.

EXPERIMENT 2

In Experiment 2, phonetic similarity of consonants to be exchanged and of their vocalic environment is controlled across conditions in which subjects exchange initial and final consonants in a syllable. In the experiment, I ask first whether the initial/final consonant difference in favor of initial exchanges is preserved when phonetic similarity of the consonants and their environments is controlled, and second whether the latency difference between the conditions can be characterized as one of differential cohesion of initial and final consonants with a syllable's vowel.

I examine the second issue by looking at errors that subjects make under instructions to exchange initial consonants, vowels, or final consonants. If initial consonants cohere with vowels less than do final consonants, as evidence reviewed in the introduction might predict, then, for example, when subjects are asked to exchange vowels, they should relatively more frequently erroneously move the final consonant with the vowel than move the initial consonant with the vowel. Similarly, when initial consonants are to be exchanged, concurrent vowel exchanges

should be rare among errors; when final consonants are to be exchanged, concurrent vowels exchanges should be relatively more common.

Method

Subjects. Subjects were 12 undergraduates at Dartmouth College who participated in the experiment for course credit. They were native speakers of English who reported normal speech and hearing.

Materials. Stimuli were 48 pairs of words; they are listed in Appendix 2. The same stimulus pairs were used in conditions in which ICs, Vs, and FCs were to be exchanged. In all conditions, both stimulus and response items were real words of English (or, in two cases, they were names). Moreover, the stimulus items were selected so that across conditions, both the stimulus items and the response items were the same. (For example, "PUN TICK" was the response for "TON PICK" in the IC-exchange condition, it was the response for "PIN TUCK" in the V-exchange condition, and it was the response for "PUCK TIN" in the exchange vowels condition.) This was accomplished by identifying 12 sets of word pairs (for example, TUCK PIN) such that permutations of initial consonants, vowels, and final consonants all created real words of the language (PUCK TIN, TICK PUN, TON PICK). Because all four permutations were used as stimulus strings in all three exchange conditions, stimulus and response items were identical across conditions; only the pairing of stimulus and response strings differed across conditions. Table 2 illustrates how this worked for the set of response strings for the family of items including "PUN TICK." By creating stimulus items in this way, the frequency of stimulus items and response items as well as the effects of response-item initial consonants on voice key were equated across experimental conditions. In addition, the similarity of the vocalic environments of the exchanging consonants was the same for initial and final consonant exchanges.

TABLE 2

Sample Stimulus and Response Pairs Used in Experiment 2.

Stimulus Strings	*Response Strings*		
	IC	V	FC
TUCK PIN	PUCK TIN	TICK PUN	TON PICK
PUCK TIN	TUCK PIN	PICK TON	PUN TICK
TICK PUN	PICK TON	TUCK PIN	TIN PUCK
PICK TON	TICK PUN	PUCK TIN	PIN TUCK

Across the items, initial and final consonants were matched in featural similarity according to the feature system of van den Broecke and Goldstein (1980). Initial consonants within a pair differed by 1.5 features on average; final consonants differed by 1.67 features.

Procedure. The procedure was essentially the same as that in Experiment 1. Exceptions were that subjects participated in three sets of trials, each consisting of one practice block and four test blocks of 12 trials each. In one set of trials, subjects exchanged initial consonants, in another, vowels, and in a third, final consonants. The order in which these sets of trials was presented was counterbalanced across the 12 subjects. As in Experiment 1, subjects' responses were recorded on audio tape. However, due to experimenter error, one subject's responses were not recorded.

Design. The experiment had one independent variable, exchanged segment, with three levels, IC, V, FC. Dependent measures were latency and percentage of errors.

Results and Discussion

Latencies and errors. Response times and error percentages are presented in Table 3. The table shows that latencies were fastest for initial consonant exchanges and slowest for vowel exchanges. For purposes of analysis, latencies were transformed into their reciprocals to correct for inhomogeneity of variance across the IC, V, and FC conditions. In an analysis of variance on the reciprocals, the main effect of exchanged segment (IC, V, FC) was highly significant, $F(2,22) = 35.27$, $p < .0001$. Post hoc analyses showed that all pairwise differences were significant.

The analysis of errors also revealed a main effect of exchanged segment, $F(2,22) = 12.87$, $p = .0002$, with the significant effect in this case due to a smaller error rate on initial-consonant exchanges than on vowel and final-consonant exchanges, which did not differ.

TABLE 3

Response Times and Percentages of Error for the Conditions of Experiment 2.

	Type of Exchange		
	IC	V	FC
RT	1606	31337	2448
Error	9.3	22.5	23.2

Analysis of errors. A major purpose of the experiment was to examine errors in which the subject moved more than just the designated segment. The question was whether whole-rhyme (henceforth, VFC) exchanges occurred disproportionately among errors of this type. To answer this question, errors from the 11 subjects whose responses had been recorded were classified into categories. In the IC exchange condition, errors were classified as ICV errors, in which the vowel was exchanged with the initial consonant, as ICFC errors in which the final consonant was moved with the initial consonant, or as "other" errors including everything else. To enable comparison with errors in the V and FC conditions, which had much higher error rates, the total errors in the ICV and ICFC conditions were each expressed as proportions of the total errors on initial consonants (that is, as conditional probabilities of an ICV or ICFC error given that an initial-consonant exchange occurred). Similarly, errors in the V condition were classifed as ICV, VFC, or as other errors and were expressed as proportions of the total errors on vowels. Next, errors in the FC condition were classified as ICFC or VFC or as other errors and were converted to proportions of total FC errors. Finally, error percentages were collapsed across conditions so that, for example, the ICV percentage contributed by the IC exchange condition and that contributed by the V exchange condition were averaged. This was done because error percentages were very closely matched across these symmetrical conditions.

Error percentages were 10.8, 36.2, and 5.5 in the categories ICV, VFC, and ICFC, respectively. As expected, if segments in the rhyme are more cohesive than segments in different syllable constituents, errors occurred disproportionately in the VFC category. Indeed, over one-third of all V and FC exchange errors were whole-rhyme exchanges. An analysis of variance on the three percentages showed a main effect of error category, $F(2,20) = 20.25$, $p < .001$. Post hoc tests showed that the difference

between the ICV and VFC categories was significant, $F(2,20) = 12.09$, $p < .0004$, while that between the two cross-constituent categories (ICV and ICFC) was not.

Post hoc analyses. The foregoing analysis shows that phoneme errors exhibit the same evidence for an internal syllabic constituent structure that other evidence from linguistic theory, linguistic games, and poetic devices suggests. In itself, however, it does not reveal any basis for the constituent structure or any hints as to why syllables partition as they do. Here I look at two possible correlates of the onset-rhyme structure that may offer deeper insight into its origins or functions, if any.

One correlate relates to the relative frequencies of initial and final consonants. For example, CVs are popular syllable types both across languages (e.g., Clements & Keyser, 1983) and in English, where, in general, words syllabify so as to maximize syllable onsets (e.g., Hoard, 1971). Possibly, syllable-initial consonants have connections to many more words in the lexicon on average than do final consonants and, thereby, are more detachable from any given word. This account might be tested by correlating consonant frequency in initial and final positions of a word with latencies to make initial and final exchanges.

For purposes of this analysis, two frequency counts were used, those of Hultzen, Allen, and Miron (1964) and of Roberts (1965). These tables provide word-initial and -final phoneme frequencies based on corpora of spoken English. Both counts are based on multi-syllabic words as well as monosyllables, so the frequency counts may not be entirely accurate for present purposes. If they are accurate, however, they show that the initial/final difference in exchange latency does not derive from a frequency difference favoring initial consonants for the stimuli in Experiment 2. Instead, for those stimuli, word-final consonants have a higher overall frequency than word-initial consonants. Therefore, correlations between consonant frequency and exchange latency are overall positive ($r = .33$, $p < .05$ for the relevant consonant in the first word of a stimulus pair and $r = .27$, $p < .05$ for the relevant consonant in the second word) for the initial- and final-consonant exchange conditions of Experiment 2.[2] However, with effects of the dichotomous variable, initial versus final consonant exchange, partialed out, the correlations are both nonsignificant. The dichotomous variable by itself, with effects of frequency partialed out, correlates strongly with latency ($r = .75$). Although the analysis does show a relationship between phoneme frequency and latency, the direction of the correlation does not support the hypothesis that high frequency of association to words renders a consonant easily detachable from a given word.

A different way of looking at frequency as a basis for cohesion is suggested by these positive correlations. Possibly, frequent cooccurrences of particular consonant and vowel pairs in words promotes cohesiveness between them. Accordingly, perhaps CV and VC diphone frequencies will be found to predict exchange latency. For these frequencies to predict the response time difference in latency between initial and final consonant exchanges, CV diphones in the stimulus pairs of Experiment 2 would have to be lower in frequency than VC diphones, and indeed there is a small difference in their frequency in the expected direction. Correlations between exchange latency and diphone frequency in *stimulus* pairs should be positive such that high diphone frequency retards detachment of an IC or an FC from its vowel. By the same reasoning, if diphone frequencies of *response* pairs predict exchange latency, correlations between the variables should be negative such that high frequencies of the newly-created CV or VC pairs promote the exchange.

For purposes of examining the effects of diphone frequency on exchange latency, Roberts' frequency counts are not appropriate, because they do not provide a fine enough transcription of the vowels in the corpus. (Roberts transcribes the vowels in the corpus as any of eight monophthongs.) Accordingly, correlatons are based on word-initial CV diphones and word-final VC diphones from the tables of Hultzen et

al. These were checked using the frequency tables of Carterette and Jones (1974). The two sets of correlations were very similar, and so just those from Hultzen et al. are reported here.

Interestingly, the correlations patterned as predicted only for the first stimulus word and the first response word of a pair--that is, for the word whose utterance by the subject triggered the voice key. For the first stimulus word, the correlation with latency is weak but positive, $r = .19$, $p = .05$; for the second stimulus word, the correlation is negative, $r = -.35$, $p < .05$. These correlations remain significant with effects of the dichotomous variable, initial- or final-consonant exchange, partialed out. For the first response word, the correlation with latency is negative, $r = -.24$, $p < .05$; for the second, it is near zero, $r = -.02$. Both of the correlations approach zero with the effect of consonant position partialed out. In a multiple regression analysis with the diphone frequencies of the first stimulus word and first response word as predictors of latency, a multiple R of .3 is obtained. This is much smaller than the correlation of .79 between the dichotomous variable, consonant position, and latency with effects of the four diphone frequency variables (first and second stimulus and response words) partialed out.

Thus, the present analysis gives only weak support to the idea that frequency affects detachability of a consonant from its vowel. For the word whose utterance triggers the voice key, high frequency of occurrence of the to-be-detached consonant and its vowel lengthens exchange latency, while high frequency of occurrence of the to-be-attached consonant and the same vowel shortens latency. However, the variable of consonant position itself explains much more of the variance in response latencies than do the diphone frequency variables. This suggests at least that different frequencies of occurrence of initial and final consonants with their vowel does not exhaust the reasons why final consonants are more difficult to detach from a syllable than are initial consonants. However, for this analysis as for the first, these conclusions must be qualified, because the frequencies used in the analysis may not be sufficiently close to those in the language user's "mental lexicon."

A different approach to examination of the effect of consonant position in the syllable on exchange latency focuses specifically on the finding that initial and final consonants appear to cohere differentially with the vowel. In particular, it relates to the possibility that that the gestural relationship between vowels and consonants is different for syllable-initial and -final consonants.

In English (Fowler, 1983), as in some other languages including Swedish (Lindblom & Rapp, 1973), a vowel in a syllable is durationally shorter the more consonants surround it in the syllable. Moreover, the shortening is asymmetrical so that syllable-final consonants shorten the vowel more than syllable-initial consonants. One account of this shortening asymmetry (Fowler, 1983) is that it is a reflection of coarticulatory overlap between consonants and vowels (e.g. Öhman, 1966; Carney & Moll, 1971). Indeed, Lindblom, Lubker, Gay, Lyberg, Branderud, and Holmgren (in press) have evidence that when closure duration of a consonant is shortened by a bite block, which enforces an unusual and constant amount of jaw opening, measured vowel duration increases by the amount of closure shortening. This would be expected if the shortening is due to a coarticulatory overlaying of the vowel by the consonant. The asymmetry in shortening presumably signifies that gestural overlap of vowels and consonants is more extensive for vowels and final consonants than for vowels and initial consonants. Of course, this difference in itself may be yet another reflection of a more fundamental onset-rhyme constituent structure to the syllable. However it may indicate at least that the reasons for the VC cohesion noticeable in errors and exchange latencies may be the same reasons, whatever they may be, for the asymmetry in gestural overlap of initial and final consonants with vowels.

A spatial or articulatory, rather than purely temporal, manifestation of the same consonant-vowel overlap might be greater cohesiveness between a vowel and a consonant on either side of it to the extent that the consonant shares articulatory or featural properties with the vowel. That is, a vowel may cohere more with consonants that share gestural or featural properties with it than with less similar consonants. If so, then the difference in the cohesiveness of initial and final consonants with the vowel might have as its basis a difference in the extent to which the production of initial and final consonants can merge with production of the vowel.

To investigate this idea, I looked at the difference in the latencies of initial and final consonant exchanges as a function of three levels of "sonority" of the consonant pairs being exchanged. "Sonority" refers to the "loudness [of a sound] relative to other sounds with the same length, stress and pitch." (Ladefoged, 1982, p. 221). Vowels are the most sonorous segments, followed in order by /l/ and /r/, nasals, fricatives and stops (e.g., Ewan, 1982). Among consonants, then, degree of sonority relates roughly to degree of articulatory vowel-likeness.

In many languages, syllable structure respects a sonority hierarchy such that consonantal segments in a cluster increase in sonority nearer the vowel. Accordingly, for example, in a prevocalic cluster containing /t/ and /r/, the order must be /tr/; but the ordering is reversed for clusters after the vowel. One way to conceptualize the difference in detachability of initial and final consonants from a vowel, compatible with the observation that segments more cohesive with the vowel occupy slots closer to it, would be to think of final consonants as occupying slots relatively closer to the vowel than do initial consonants.

Although in Experiment 2, I controlled overall phonetic similarity of initial and final consonants, I did not control degree of sonority. To examine any role that sonority might play in the degree of cohesiveness between vowel and consonant, I partitioned stimulus pairs into three categories according to the sonority of the consonants that were exchanged. In one category exchanged consonants were stops; in a second category one member of a exchanged pair was a stop and one was some more sonorous consonant; in the third category, neither segment of a pair was a stop. Table 4 provides the average latencies of exchanges for stimuli in these categories, distinguished by the initial or final position of the consonant pair in the word.

TABLE 4

Latencies of Initial and Final Exchanges Distinguished by the Sonority Exchanged Consonants.

	SONORITY		
	Low	Middle	High
Initial	1534	1590	1692
Final	2333	2431	2615

The results show an increase in latency with consonantal sonority among both initial and final consonants. This is consistent with the idea that more vowel-like consonants are less detachable from the vowel than others. However, with initial and final consonants essentially matched in sonority (for example, in trials where consonants being exchanged are uniformly stops), there remains a very large difference in latency between initial and final consonant exchanges, suggesting strongly that a tendency for initial and final consonants to differ in sonority does

not go far to explain the difference between them in exchange latency and, presumably, also in cohesiveness with the vowel.

An analysis of variance performed on reciprocals of the latency data summarized in Table 4 revealed significant effects of consonant position, $F(1,90) = 214.24$, $p < .0001$, and of sonority, $F(2,90) = 3.50$, $p = .03$, but no interaction between the variables, $F<1$.

In summary, then, the analysis does suggest that more vowel-like consonants are harder to detach from the vowel than less vowel-like consonants. However, this difference by itself cannot explain the effects of consonant position on exchange latency, because with sonority of ICs and FCs essentially matched, large differences in latency remain. Of course, it is possible, even likely, that final consonants and vowels do overlap more in production than do initial consonants and vowels, perhaps because as just suggested, in some sense, final consonants occupy slots closer to the vowel than do initial consonants.

A speculative reason why final consonants may nestle closer to the vowel derives once again from considerations of temporal coarticulatory overlap in a CVC syllable. The articulation of a consonant may be characterized as including three phases, a closing phase in which its characteristic constriction is approached, a closure phase in which the constriction is maintained, and a release into a following segment, if any. A vowel need have only an opening phase; in addition, in slow speech, it may have a phase in which the vocal-tract shape that is the target of the opening gesture is maintained.

Coarticulatory overlap between the initial consonant and the primary articulations for a vowel (that is, articulations involving the tongue body and jaw that produce the characteristically open tract shape for the vowel), occurs when vowel production is initiated during consonant closure or release. The early phases of vowel production will consist of positioning movements of the tongue body and of jaw opening. The opening gesture cannot occur too early within the initial consonant, else it will cause premature release of the consonant and an acoustic signal that does not specify the intended consonantal phoneme. Possibly, then, temporal overlap between the initial consonant and the jaw-opening gestures for the vowel is constrained by the gestural requirements of the consonant. Of course, as other theorists have noted (e.g., Daniloff & Hammarberg, 1973), some overlap between consonant and vowel is essential to prevent production of unintended vocalic sounds on release of the consonant.

On the other side of the vowel, overlap with the final consonant occurs when the jaw and other primary articulators for the consonant begin the consonant's closing phase. On this side, more overlap between vowel and consonant may be possible, because the final consonant has only to permit enough opening for the vowel so that the vowel's identity is specified in the acoustic signal.

Accordingly, pehaps in the planning of a syllable or monosyllabic word where phoneme errors appear to arise (e.g., Dell, 1980), the possibilities for greater V-FC than IC-V overlap are reflected in a way that gives rise not only to asymmetrical coarticulatory overlap and shortening, but also to differences in initial- and final-consonant exchange frequencies and to the other manifestations ascribed to an onset-rhyme structure of the syllable. That "way" is, indeed, what the claim that syllables have an onset-rhyme constitutency may represent.

ACKNOWLEDGMENT

The research reported here was supported in part by NIH Grant HD 01994 to Haskins Laboratories. I thank Kristin Snow for helping with data collection and analysis and George Wolford for commenting on a draft of the manuscript.

REFERENCES

Baars, B. J., & Motley, M. (1976). Spoonerisms as sequencer conflicts: Evidence from artificially elicited errors. *American Journal of Psychology, 89*, 467-474.

Baars, B. J., Motley, M., & MacKay, D. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382-391.

Carney, P., & Moll, K. (1971). A cinefluorographic investigation of fricative-consonant vowel coarticulation. *Phonetica, 23*, 193-201.

Carter, R., & Bradshaw, J. (1984). Producing 'Spoonerisms' on demand: Lexical, phonological and orthographic factors in a new experimental paradigm. *Speech Communication, 3*, 347-360.

Carterette, E., & Jones, M. (1974). *Informal speech*. Berkeley, CA: University of California Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Cutler, A. (1981). The reliability of speech error data. *Linguistics, 9*, 561-582.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics, 1*, 239-248.

Dell, G. (1980). *Phonological and lexical encoding in speech production: An analysis of naturally occurring and experimentally-elicited speech errors*. Unpublished doctoral dissertation, University of Toronto.

Dell, G. (1986). A spreading-activation theory of retrieval in speech production. *Psychological Review, 93*, 283-321.

Dell, G., & Reich, P. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611-629.

Ewan, C. (1982). The internal structure of complex segments. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. 2, pp. 27-68). Dordrecht, The Netherlands: Foris Publications.

Fowler, C. A. (1983). Converging sources of evidence for spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General, 112*, 386-412.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton-Mifflin.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Garrett, M. (1975). The analysis of speech production. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9). New York: Academic Press.

Garrett, M. (1976). Syntactic processes in sentence production. In R. J. Wales & E. C. Walker (Eds.), *New approaches to language mechanisms* (pp. 231-255). Amsterdam: North Holland.

Garrett, M. (1980). Levels of processing in speech production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177-220). London: Academic Press.

Hoard, J. (1971). Aspiration, tenseness and syllabication in English. Language, 47, 133-140.

Hultzen, L., Allen, J., & Miron, M. (1964). *Tables of transitional frequencies of English phonemes*. Urbana: University of Illinois Press.

Kupin, J. (1979). *Tongue twisters as a source of information about speech production*. Doctoral dissertation, University of Connecticut.

Ladefoged, P. (1982). *A course in phonetics* (2nd ed.). New York: Harcourt Brace Jovanovich.

Levitt, A., & Healy, A. (1985). The roles of phoneme frequency, similarity and availability in the experimental elicitaion of speech errors. *Journal of Memory and Language, 24*, 717-733.

Lindblom, B., Lubker, J., Gay, T., Lyberg, B., Branderud, P., & Holmgren, K. (in press). The concept of target and speech timing. In R. Channon & L. Shockey (Eds.), *In honor of Ilse Lehiste*. Providence, RI: Foris.

Lindblom, B., & Rapp, K. (1973). Some regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm, 21*, 1-59.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-67.

MacKay, D. G. (1971). Stress pre-entry in motor systems. *American Journal of Psychology, 84*, 35-51.

MacKay, D. G. (1978). Speech errors inside the syllable. In A. Bell & J. Hooper (Eds.), *Syllables and segments* (pp. 201-213). Amsterdam: North Holland.

Motley, M., & Baars, B. (1975). Encoding sensitivity to phonemic markedness and transitional probability: Evidence from Spoonerisms. *Human Communication Research, 2*, 351-361.

Öhman, S. (1966). Coarticulation in VCV syllables; Spectrographic measurements. *Journal of the Acoustical Society of America, 39,* 151-168.

Prince, A. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry, 11,* 511-562.

Roberts, A. H. (1965). *A statistical analysis of American English.* The Hague: Mouton.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. Walker (Eds.), *Sentence production: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech.* New York: Springer-Verlag.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech-error data. *Journal of Verbal Learning and Verbal Behavior, 18,* 41-55.

Treiman, R. (1985). Onsets and rimes as units in spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 35,* 161-181.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25,* 476-491.

van den Broecke, M., & Goldstein, L. (1980). Consonant features in speech errors. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 47-66). New York: Academic Press.

FOOTNOTES

**Speeech Communication.*

†Also Dartmouth College.

[1]The conclusion that there is a lexical bias here is weakened by the presence of a kind of bias in the stimuli. Two thirds of the response strings form words; accordingly, subjects may have come to expect word outcomes. However, a similar imbalance in stimulus strings did not lead to a compatible advantage for stimulus strings that form words.

[2]These correlations are based on the analysis using tables from Hultzen et al. Those based on Roberts' tables were similar in magnitude and pattern. Subsituting log frequencies for raw frequencies did not improve the correlations, and so correlations involving raw frequencies are reported throughout.

APPENDIX 1

PHONETICALLY SIMILAR:

W-W	W-NW	NW-W
barn door	bars dorm	bart dord
lake room	lame rule	lail roop
shale sun	shade sup	shafe sutt
gin choice	gist choate	jick choe
paste toll	pain toad	pame tope
cat pane	cab paid	cal pake
zip sue	zig soup	zill soom
ball gun	bought gull	bolf guck
wield yell	week yes	weast yed
tall cone	toss cope	toff coes
mutt nail	muss nap	mub nake
feel poke	feast pope	feep pome

PHONETICALLY DISSIMILAR:

barn your	bars yawn	baht yack
lake cone	laze coast	lape cobe
shale pone	shame pour	shane pode
gin sam	gym sad	jick soe
pest gain	pen gaze	pess gare
cat sum	calf sun	cack suff
zip new	zig newt	zill soom
balk chaste	bought choir	baw chail
wield poke	wean pole	weel pore
toll route	tall rum	toach rick
mat soul	man soak	mand soat
feel done	feast dub	fean dus

APPENDIX 2

root shod	rude shot	rot shooed	rod shoot
meal not	meat knoll	mole neat	moat kneel
feet made	feed mate	fate mead	fade meet
bait well	bail wet	bet wail	bell wait
sip hole	sill hope	soap hill	soul hip
moan bad	mode ban	man bode	mad bone
feel put	feet pull	full peat	foot peel
top pick	tock pip	tip pock	tick pop
rang sum	ram sung	rung Sam	rum sang
date roll	dale rote	dote rail	dole rate
tuck pin	ton pick	tick pun	tip puck
duel tine	dune tile	dial tune	dine tool

Word-level Coarticulation and Shortening in Italian and English Speech*

Mario Vayra,† Carol A. Fowler,†† and Cinzia Avesani†

Our study compares measures of word-level coarticulation and shortening in Italian and English. In English, these measures have been shown to correlate and to mirror the supposed left-dominant foot structure of English words. That is, stressed syllables coarticulate more with, and are shortened more by, unstressed syllables that follow them than unstressed syllables that precede them. A recent study of Italian word stress (Nespor & Vogel, 1979) suggests that, like English, Italian has a left-dominant foot structure. Our investigation of coarticulation and shortening in Italian shows no evidence that this foot structure is realized in the coarticulation or timing behavior of talkers. Three Italian talkers differed markedly from each other and from English talkers on measures of coarticulation and shortening.

INTRODUCTION

In the present study we examine patterns of within-word vowel-to-vowel coarticulation and shortening in Italian and compare the patterns with those found in similar English productions. The comparison is of interest in light of evidence linking coarticulatory and durational patterns of English to the rhythmical character of the language.

English is identified as stress timed, having strong and weak syllables that approximately alternate and, by some accounts (e.g., Abercrombie, 1964; Pike, 1945), having a tendency for strong or stressed syllables to be evenly timed. Inter-stress intervals in English are not isochronous; instead they vary in duration with several variables, including the phonetic composition of their component syllables and, especially, the number of syllables in the interval (e.g., Classe, 1939; Shen & Peterson, 1962).

A tendency toward stress timing is identified in several ways. Jassem, Hill, and Witten (e.g., 1984) report that the slope of a line relating the normalized interval between stresses to its normalized number of constituent phonemes is less than one. That is, the more phonemes, and presumably syllables, in an interval between stresses, the shorter the average phoneme duration in the interval. Although Jassem et al. interpret this in favor of a stress-timing tendency in English, it could, instead, be the consequence of the general fact that stressed syllables are longer than unstressed syllables; the more segments in an interstress interval, in general, the larger the ratio of unstressed to stressed segments and, consequently, the shorter the average segment duration. More telling, perhaps, is a finding that stressed vowels shorten in the context of following unstressed syllables (e.g., Fowler, 1981). This

effect is also seen in Swedish (Lindblom & Rapp, 1973), Dutch (Nooteboom, 1973), and German (Kohler, 1983), languages also identified as stress timed. Interestingly, in these languages, where a comparison has been made, stressed syllables shorten little in the context of preceding unstressed syllables. This suggests an organization of syllables in which stressed syllables are grouped with following, but not with preceding, unstressed syllables.

This same conclusion has been drawn independently by investigators attempting to characterize systematic stress placement in English words and sentences (e.g., Selkirk, 1980). They identify sequences, called "feet," that include a stressed syllable and following unstressed syllables as basic rhythmic constituents of an utterance.

A third source of evidence for a stress-timing tendency in English is that some unstressed vowels reduce to schwa. This enables additional shrinkage in feet containing unstressed syllables.

Finally, English is supposedly subject to a "Rhythm Rule" or, more generally, to rules of "eurhythmy" (see Hayes, 1984; but see Cooper & Eady, 1986), that serve to adjust lexical stress in a word in the context of other words so as to keep strong syllables in approximate alternation with weak ones. (So, for example, English speakers say "Tennes*see*" and "Tennes*see* in a nutshell," with a strong final syllable, but "*Ten*nessee Ernie" with the stress shifted back to the first syllable, away from the strong first syllable of "Ernie.")

Coarticulatory patterns in English appear to reflect the foot structure of the language. In particular, the influence of stressed vowels on following (trans-consonantal) unstressed vowels is strong while their influence on preceding unstressed vowels is weak (Fowler, 1981). This pattern matches the shortening asymmetry mentioned previously whereby a stressed syllable is shortened substantially by following unstressed syllables and only weakly by preceding ones. Both sets of findings suggest an asymmetrical cohesion between stressed syllables and following unstressed syllables on the one hand, and stressed syllables and preceding unstressed syllables on the other. Indeed, one study (Fowler, 1981) found that measures of coarticulation and shortening are correlated in English talkers such that a stressed vowel that has a marked coarticulatory influence on a syllable is shortened by that syllable substantially, whereas one that exerts a weak coarticulatory effect on a neighbor is shortened by the neighbor very little. Fowler suggested that, in English, vowel-to-vowel coarticulation and shortening may be alternative measures of the same tendency for production of stressed vowels to be overlaid by the production of following unstressed vowels largely within a foot.

For its part, Italian is identified as syllable-timed, not stress-timed. The evidence favoring this classification is equivocal, however. Using the analysis of Jassem et al., Bertinetto (1981a, 1981b) found very little evidence of compression in normalized intervals between stresses as a function of their normalized size in either segments or syllables. This contrasts with the findings of Jassem et al. on speakers of English; whereas Jassem et al. had found a slope less than one relating normalized inter-stress interval to normalized interval size in number of segments, Bertinetto found a slope of nearly one.

However, other studies have found contradictory evidence. A large duration difference between stressed and unstressed vowels in speakers of Italian as in speakers of Dutch, a stress-timed language, is reported by den Os (1985). Other things equal, this would give a slope less than one between normalized inter-stress interval and normalized interval size in the den Os data.

In addition, Farnetani and Kori (1984) and Vayra, Avesani, and Fowler (1984) both report shortening of stressed vowels due to following unstressed syllables in a word. In the latter study, shortening in the speech of a single Italian speaker was substantially weaker than in comparable productions of English talkers. Vayra et al. also found some very limited shortening due to preceding unstressed vowels, whereas

Farnetani and Kori (1983) found none at all (but see Farnetani & Kori, 1982). Marotta (1984) found weak shortening of a stressed syllable followed by two as compared to one unstressed syllables. This shortening was absent in a four syllable word with first syllable stress, however. These findings hint at a foot structure similar to that in English, but that is less strongly marked in Italian than in English segment durations.

Compatible with this pair of findings is some preliminary work on word stress in Italian by Nespor and Vogel (1979). They propose that word stress in Italian is best accounted for by assuming that the language has a foot structure consisting of a stressed syllable and following unstressed syllables. Unstressed syllables preceding a stressed syllable are grouped with the preceding foot. These same investigators also propose that some dialects of Italian exhibit a rhythm rule similar to English. (However, Bertinetto, 1985, has recently questioned whether a rhythm rule similar to that operating in English plays a major role in Italian phonology.)

As for a syllable-timing tendency in Italian, Bertinetto (1981a, 1981b) found the same slope of one relating syllable duration to syllable size in segments as he had found for intervals between stresses. However, both Vayra et al. (1984) and Farnetani and Kori (1984) report evidence of some shortening of a vowel in the presence of neighboring consonants in the same syllable. The shortening is asymmetrical, with shortening due to postvocalic consonants greater than that due to prevocalic consonants. This asymmetry itself, of course, is not predicted by a syllable-timing constraint.

Just as Italian shows intimations of stress timing, so English shows shortening of vowels due to consonants, just given as evidence of syllable timing. In English, (Fowler, 1983; see also Lindblom & Rapp, 1973, for Swedish), vowels are shortened by consonants in the same syllable; as in Italian, in English and Swedish vowels are shortened more by following than by preceding consonants.

This collection of findings leads to the possibility that stress timing and syllable timing are not mutually exclusive timing types (see Dauer, 1983, for a similar conclusion). Rather, each may represent a level of a tiered set of timing or rhythmical constraints that languages share (cf. Liberman & Prince, 1977). However, languages differentially emphasize one level over others, or, in the case of Italian perhaps, emphasize no level particularly.

In the present study, we examine measures of coarticulation and shortening in Italian and compare them directly to the same measures in English to ask whether the measures pattern in similar ways in the two languages, and in particular, in ways that can be identified with a foot structure and stress timing. If Italian has a foot structure similar to English, as Nespor and Vogel suggest, and if coarticulatory and shortening patterns reflect that structure, then we would expect Italian speakers to show an asymmetry in coarticulatory influences with carryover influences of stressed on unstressed vowels stronger than anticipatory influences and to show more shortening of a stressed vowel due to following than preceding unstressed syllables. However, as noted, there are indications that this pattern, if present, should be weaker in Italian than in English. Stress timing in Italian is not salient enough for researchers to have identified the language as stress timed. Accordingly, rhythmical tendencies at the level of the foot may be weak.

Compatibly, as noted, the shortening reported by Vayra et al. (1984) is weak compared to that found in English. Indeed, Bertinetto (1981a, 1981b) found even less evidence for it. Therefore, if present, the durational and coarticulation asymmetries in Italian may be less marked than they are in English. We may also predict that Italian speakers will show less coarticulatory influence of stressed vowels on unstressed neighbors overall, because, lacking the malleable schwa, the language may permit less coarticulatory influence on its unstressed vowels.

Methods

Subjects

Italian talkers were three Piedmontese (Northwestern Italy) speakers of Standard Italian. Their speech has some features characteristic of Northern pronunciation.[1] All three speakers have a university education; two (PMB, MC) teach and the third (CB) is a researcher at the university level. English speakers were two speakers of American English (Northeastern dialects). Both have university educations; one is the second author.

Stimuli

Tables 1 and 2 present the target real- and pseudo-words produced respectively by the Italian and English speakers. Words were one, two, three, or four syllables in length, with the position of the stressed syllable or syllables manipulated. Each pseudoword was matched to a real word with which it shared its stress pattern.

TABLE 1[2]

Pseudowords and their real-word counterparts produced by Italian talkers in the carrier sentence, "Ora diciamo - due volte." (The symbol [ˌ] marks secondary prominence; the symbol ['] primary prominence).

fi	qua
fa	
fa'fi	pe'ro
fa'fa	
fa'fifa	cu'cina
fa'fafa	
fafa'fi	libe'ro
fafa'fa	
ˌfifa'fa	ˌdall' a 'me
ˌfafa'fi	
ˌfafa'fa	
'fifa	'cara
'fafa	
'fifafa	'tavola
'fafafa	
'fifafaˌfa	'portameˌne
'fafafaˌfi	
'fafafaˌfa	

For Italian speakers the pseudowords consisted of the stressed syllables /fi/ and /fa/ and the unstressed syllable /fa/. For English speakers, the corresponding stressed syllables were /si/ and /sʌ/, and the unstressed syllable was /sə/.[3] Two of the stress patterns we used require further comment. We used "mistletoe" as a real word template in English to represent a stress pattern in which the first and third syllables receive some stress. There are no words in Italian with such a stress pattern; therefore, we used the phrase "dall' a me" to represent that stress pattern. This phrase is normally produced with weakened prominence on the first syllable and primary stress on the last syllable (see Farnetani & Kori, 1983; Marotta, 1984, for similar evidence of "de-accenting" at the phrase level). An alternative pronunciation, however, is to stress only the last word. We used "ironingboard" in English to represent a lexical pattern with first and fourth syllable stress.[4] Once again, we could not find an analogous pattern in Italian. We used "portamene" with primary stress on

the first syllable. This word may have rhythmical enhancement of prominence on the last syllable, but that syllable is not lexically stressed (Bertinetto, 1981a).

The real and pseudo-words were produced in a carrier sentence ("Ora diciamo - due volte" [Now talk - two times] for Italian talkers; "Now talk - two times" for English talkers).

TABLE 2

Pseudowords and their real-word counterparts produced by English talkers in the carrier sentence "Now talk - two times."

si	day
sʌ	
sə'si	a'sleep
sə'sʌ	
sə'sisə	de'licious
sə'sʌsə	
səsə'si	inter'vene
səsə'sʌ	
'sisə,sʌ	'mistle,toe
'sʌsə,si	
'sʌsə,sʌ	
'sisə	'easy
'sʌsə	
'sisə sə	'ignorant
'sʌsə sə	
'sisə sə,sʌ	'ironing,board
'sʌ sə sə,si	
'sʌsə sə,sʌ	

Procedure

Pairs of sentences, one containing a real target word and the other its pseudoword counterpart, were printed on file cards. The cards were randomly ordered.

Subjects were tested individually in a quiet room. After practice reading the sentences, subjects read through the deck of file cards four times. The order of the cards in the deck was randomized after each reading, so that each subject read four different random orders. On each card, subjects read both sentences. The four readings of each card were recorded on audiotape.

Spectrographic displays provided the bases for measurements of coarticulation. We assessed the coarticulatory influence of a stressed vowel on its neighboring unstressed vowels by subtracting the center frequency of an unstressed vowel in the neighborhood of stressed /a/ (or, in English, stressed /ʌ/) from its value in the neighborhood of stressed /i/. (For example, the center frequency of F_2 of unstressed /a/ in /'fafa/ was subtracted from the center frequency of F_2 of unstressed /a/ in /'fifa/.) We refer to this measure as the "coarticulation score."

Measurements of duration were made from waveform displays of the Italian speech and from spectrographic displays of English speech. Measurements were made of the durations of stressed and unstressed vowels defined as the intervals of periodicity in each syllable.

Results

Coarticulation

Tables 3 and 4 provide cell means for Italian and English talkers on one set of analyses described below. These tables show similarities and differences between the language groups. Speakers in both groups tend to show more coarticulation from adjacent stressed vowels than from nonadjacent stressed vowels especially for carryover coarticulation. English speakers show consistently greater carryover than anticipatory coarticulation. Italian speakers as a group do not.

TABLE 3

Coarticulation scores (Hz) for the three Italian talkers.

	CB	PMB	MC	Composite	
ANTICIPATORY COARTICULATION					
fa'fi	86.8	65.3	71.5	74.5	Adjacent
fa'fifa	2.5	6.0	114.8	41.0	Adjacent
fafa'fi	41.5	214.5	78.3	111.4	Adjacent
fafa'fi	-5.8	47.8	151.8	64.5	Nonadjacent
ˌfafa'fi	41.5	-7.0	80.5	38.4	Adjacent
'fafafaˌfi	20.5	77.5	-23.5	24.8	Adjacent
'fafafaˌfi	5.8	41.5	11.3	19.5	Nonadjacent
CARRYOVER COARTICULATION					
'fifa	71.3	0	98	56.4	Adjacent
fa'fifa	47.8	71.8	98.5	72.7	Adjacent
'fifafa	107.3	47.5	203	119.3	Adjacent
'fifafa	89.3	-28.5	17.8	26.2	Nonadjacent
ˌfifa'fa	41.8	24.8	98.3	54.5	Adjacent
'fifafaˌfa	59.5	35.8	193.3	100.2	Adjacent
'fifafaˌfa	18.3	-41.8	0	-7.8	Nonadjacent

Below we describe two sets of analyses performed on the data for the Italian and English speakers. One analysis is designed to test for coarticulation itself. The second examines the change in magnitude of coarticulation with distance from the stressed vowel. In these analyses and following ones, any factor in the analysis of variance that we do not discuss explicitly is nonsignificant.

Italian talkers. Separate three-way analyses of variance with factors Context Stressed Vowel (/i/, /a/), Direction of Coarticulation (anticipatory, carryover), and Stress Pattern (see Tables 3 and 4 for the seven levels of this factor) were performed on the values of F_2 from each speaker. Individual tokens served as the random factor. An effect of context stressed vowel with F_2 of unstressed /a/ higher in the context of /i/ than of /a/ would signify an overall coarticulatory effect. All three Italian speakers show a large effect of context vowel (CB: $F(1,84) = 32.9$, $p < .0001$; PMB: $F(1,84) = 15.55$, $p = .0002$; MC: $F(1,84) = 30.16$, $p < .0001$).

A second factor of interest is the interaction of context stressed vowel with direction of coarticulation. This would be significant if the effect of context vowel,

just described, were different for preceding than for following stressed vowels—that is, if carryover and anticipatory coarticulation were different in magnitude. The interaction was significant for two talkers, but for different reasons. Talker CB had larger carryover than anticipatory coarticulation, $F(1,84) = 6.67$, $p = .01$. However, PMB showed the reverse asymmetry, $F(1,84) = 5.51$, $p = .02$. MC showed substantial coarticulation in both directions; the numerical asymmetry favored carryover coarticulation, but the difference did not approach significance.

TABLE 4
Coarticulation scores (Hz) for the three English talkers.

	ANTICIPATORY COARTICULATION			
	GD	CAF	Composite	
sə'si	83	28	56	Adjacent
sə'sisə	74	84	79	Adjacent
səsə'si	97	84	90	Adjacent
səsə'si	95	13	55	Nonadjacent
'sʌsə,si	130	98	114	Adjacent
'sʌsəsəsi	37	56	46	Adjacent
'sʌsəsə,si	14	14	14	Nonadjacent
	CARRYOVER COARTICULATION			
	GD	CAF	Composite	
'sisʌ	153	125	139	Adjacent
sə'sisə	134	153	144	Adjacent
'sisəsə	195	181	188	Adjacent
'sisəsə	51	28	39	Nonadjacent
'sisə, sʌ	186	153	169	Adjacent
'sisəsə,sʌ	199	92	145	Adjacent
'sisəsə,sʌ	93	69	82	Nonadjacent

The remaining factors in the analysis involve the factor stress pattern. Effects of this factor can be seen more clearly in a different type of analysis using the coarticulation score, previously described, rather than separate F_2 values for the /a/ and /i/ contexts. Table 3 represents the cell means of an analysis of variance using the coarticulation score as the dependent measure. To examine the effect of degree of proximity of the stressed vowel to the unstressed vowel supplying the dependent measure, we defined two degrees of distance. An unstressed syllable might be adjacent to the stressed syllable (as in the second syllable of 'fifafa) or nonadjacent (as in the third syllable of 'fifafa). In a planned comparison of the two degrees of adjacency, talkers CB and MC showed marginal adjacency effects: $F(1,42) = 2.64$, $p = .11$; $F(1,42) = 2.99$, $p = .08$, respectively). The adjacency effect was significant for PMB, $F(1,42) = 4.25$, $p = .04$. In general, then, the talkers show a tendency for coarticulatory effects to diminish with distance, but the effect is weak. In the overall analyses, the effect of stress pattern did not interact with the factor direction of coarticulation; therefore the distance effect is no stronger among syllables ostensibly within a foot (carryover coarticulation) than among syllables in different feet (anticipatory coarticulation).

English talkers. The data for the two English talkers were analyzed in the same way as those for the Italian talkers. The effect of context stressed vowel (/i/ or /ʌ/) was

significant (GD: $F(1,84) = 11.96$, $p = .001$; CF: $F(1,84) = 16.96$, $p = .0001$). In addition, the interaction of that factor with direction of coarticulation was significant for both talkers (GD: $F(1,84) = 13.12$, $p = .0006$; CF: $F(1,84) = 8.60$, $p = .0044$). For both talkers the interaction reflected stronger carryover than anticipatory coarticulation.

The effect of stress pattern was examined in a separate analysis with the coarticulation score as the dependent measure. However, because, in contrast to the Italian talkers, the English talkers displayed similar coarticulatory patterns, their adjacency effects were tested on a single overall analysis of variance with Talker, Stress Pattern and Direction of Coarticulation as factors. Both talker and token served as random factors. A planned comparison of the adjacency effect was highly significant: $F(1,6) = 50.44$, $p < .001$. This effect did not interact with any other variables.

Duration

Figures 1 and 2 plot duration measures of the target stressed /i/ vowels for Italian and English speakers, respectively. The left-most point in each panel of each figure is the duration of /i/ in a monosyllable, and next to it is /i/ flanked by unstressed syllables. The remaining target contexts can be classified as reflecting shortening due to one or two following syllables (points labeled A for "anticipatory" shortening due to following syllables; see Lindblom & Rapp, 1973) or due to one or two preceding syllables (points labeled B for "backward" shortening).

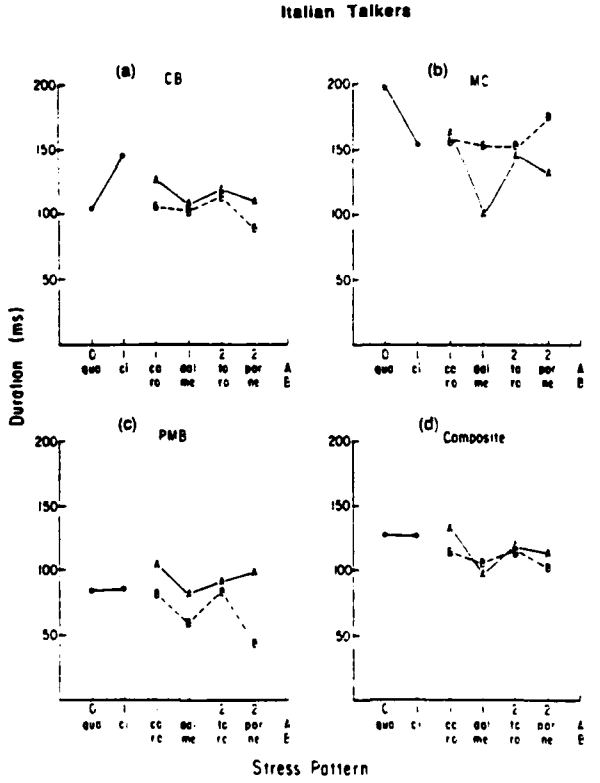


Figure 1. Durations (ms) of stressed /i/ in pseudowords spoken by Italian talkers. Points labeled "A" refer to "anticipatory shortening," due to following unstressed syllables. Points labeled "B" refer to "backward shortening," due to preceding unstressed syllables. Across the abscissa: fi, fa'fifa. A: 'fifa, ,fifa'fa, 'fifafa, 'fifafa,fa. B: fa'fi, ,fafa'fi, fafa'fi, 'fafafa,fi. Numeric labels represent the number of following (A) or preceding (B) unstressed syllables in the word. Alphabetic labels identify the stressed syllables of the corresponding template real words of Table 1.

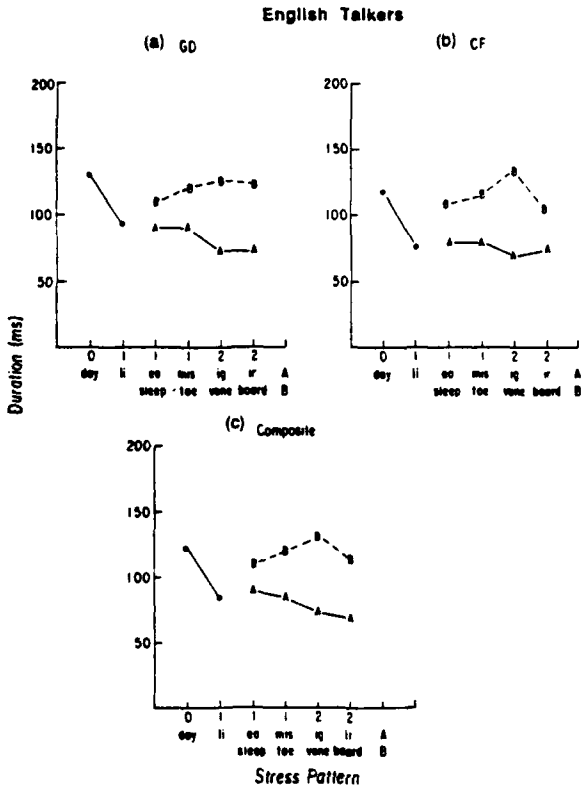


Figure 2. Durations (ms) of stressed /i/ in pseudowords spoken by English talkers. Points labeled "A" refer to "anticipatory shortening," due to following unstressed syllables; points labeled "B" refer to "backward shortening," due to preceding unstressed syllables. Across the abscissa: si, sə'sisə, A: 'sisə, 'sisəˌsʌ, 'sisəsə, 'sisəsəˌsʌ; B: sə'si, 'sʌsəˌsi, səsə'si, 'sʌsəsəˌsi. Numeric labels represent the number of following (A) or preceding (B) unstressed syllables in the word. Alphabetic labels identify the stressed syllables of the corresponding template real words of Table 2.

In a language showing stress-timing influences, the duration of /i/ should be longest in the monosyllable and shortest in the context of two rather than one neighboring unstressed syllables. In a stress-timed language with left dominant feet, this latter effect should be limited largely to contexts of following, rather than preceding, unstressed syllables.

Italian talkers. Only one of the Italian talkers fits the foregoing description in most respects. Only for talker MC is /i/ longest in the monosyllable. Also only for this talker do anticipatory shortening contexts lead consistently to shorter stressed vowels than comparable preceding shortening contexts.

Analysis of variance with factors Stress Pattern (with the ten levels plotted in the figures) and Talker, gave a significant effect of Talker, $F(2,90) = 362.3$, $p < .001$, and a significant interaction of Talker with Stress Pattern, $F(18,90) = 11.39$, $p < .001$. The interaction is significant because different talkers showed different durational patterns among the target words. Accordingly, separate one-way analyses of variance were performed on the data from each talker.

All three talkers had significant overall effects of Stress Pattern (MC: $F(9,30)=16.47$, $p < .001$; PMB: $F(9,30) = 109.5$, $p < .001$; CB: $F(9,30)= 6.43$, $p < .001$). Post

hoc (Scheffé's) tests tested various hypotheses concerning the source of the significant effect of stress pattern. We tested the difference between the four anticipatory-shortening contexts and the four backward-shortening contexts. This was significant for two talkers (MC: $F(9,30) = 3.14$, $p = .009$; PMB: $F(9,30)=5.40$, $p = .002$). Interestingly, for these two talkers, the significant effects were in opposite directions, and both talkers showed the same asymmetry in shortening as they had in coarticulation. MC had shown (numerically, not significantly) greater carryover than anticipatory coarticulation, suggesting greater cohesion between a stressed vowel and following unstressed syllables. His shortening asymmetry suggests the same asymmetry in cohesion. Both effects are consistent with a left-dominant foot structure. For his part, PMB shows greater anticipatory than carryover coarticulation, suggesting greater cohesion between stressed syllables and preceding unstressed syllables. His shortening asymmetry suggests the same cohesion. Both effects are consistent with a right-dominant foot.

CB had shown significantly more carryover than anticipatory coarticulation. His shortening effects are not consistent in direction with this. The shortening does not approach significance; however, Figure 1a reveals a numerical tendency to favor backward shortening.

It is not possible, with just three talkers, to evaluate whether the outcome shown by MC and PMB that coarticulation and shortening show the same direction of asymmetry is typical of Italian talkers or not. If the two possible asymmetries in durational shortening (ignoring the possibility of symmetry) were randomly paired with the two possible asymmetries in coarticulation, four pairings would result of which three are represented by the three talkers in the study.

In another set of post hoc tests, we asked whether two unstressed syllables led to more shortening than one. No talker showed significance in these tests.

English talkers. As for the coarticulation findings, durational patterns were more similar for the two English talkers than for any pair of Italian talkers.

An analysis of variance with factors Talker and Stress Pattern showed significant main effects (Talker: $F(1,60) = 10.85$, $p = .0018$; Stress Pattern: $F(9,9) = 30.10$, $p < .001$), but no interaction, $F(9,60) = 1.03$. Thus although the talkers differed in the overall duration of the stressed vowels they produced, the pattern of durations among the target words was the same for both talkers.

Separate analyses performed on the data from each talker verified that the effect of stress pattern was significant for both talkers individually. Post-hoc tests found anticipatory shortening more extensive than backward shortening (GD: $F(9,30)=12.17$, $p < .001$; CF: $F(9,30)=10.52$, $p < .001$). Two following unstressed syllables lead to numerically more shortening of stressed /i/ than one (GD: 90 ms versus 74 ms; CF: 80 ms versus 68 ms), but the differences are nonsignificant. Two preceding unstressed syllables do not shorten a stressed /i/ more than one. Indeed, there is little or no backward shortening in the data.

The data from these talkers conform well to expectations based on the foot structure of English, the coarticulatory patterns shown by the talkers and with one exception, they also conform well with the hypothesis (Fowler, 1981) that both measures, shortening and coarticulation, redundantly reflect the coproduction of stressed with unstressed vowels in English. The exception is that, whereas both talkers exhibit consistent anticipatory coarticulation, backward shortening, at least in the data of CF, is absent. This may reflect a difference in the sensitivity of the two measures, or it may, in fact, reflect an indication that coarticulation need not imply concomitant shortening.

The Relation between Coarticulation and Shortening

Under a hypothesis that coarticulation is overlapping production of neighboring segments, (Fowler, 1981, 1983), durational shortening is expected (other things equal) as a natural consequence of coarticulation. That is, because a coarticulating unstressed syllable "covers over" an edge of a stressed syllable, the stressed syllable is measured as shorter.

Among the six English talkers examined by Fowler (1981), five showed strong relationships betwen coarticulation and shortening such that segments coarticulating strongly with their neighbors were shortened substantially by them.

Here we attempt an analogous analysis of the Italian and English data. In the earlier research, we had used vowels produced in a stressed monosyllable as a standard "unshortened vowel" against which shortening could be assessed. This was not possible in the present analysis because only one of the Italian talkers had vowels in monosyllables longer than those elsewhere. Accordingly, we compared the duration itself of a stressed vowel with an assessment of its coarticulatory influence. In this type of analysis, the expected correlation is negative such that more coarticulation is characteristic of a shorter stressed vowel.

A problem arises in words with two unstressed vowels. In these words, there are two measures of coarticulation (one on each unstressed vowel) but just one measure of stressed vowel duration. We elected to sum the two measures of coarticulation to reflect the extended influence (if any) of stressed vowels in these contexts. This may inflate the correlation, but it appears a more appropriate measure than averaging the effects or ignoring one of them. In any case, our aim is to compare talkers of Italian and English, and the same inflationary tendency is present for speakers of both languages.

For the two English talkers, the correlation was significant and negative (GD: $r = -.89$, CF: $r = -.79$). It was nonsignificant for all three Italian talkers. For MC and PMB, who, as already noted, had durational and coarticulatory asymmetries that were consistent in direction (but opposite for the two talkers), correlations were negative (MC: $r = -.25$; PMB: $r = -.34$). For CB, the correlation was positive ($r = .14$).

Durations of Stressed and Unstressed Syllables in Different Positions in the Word

Farnetani and Kori (1982) report little variation in durations of unstressed syllables in different contexts of word length and position in the word. We examined durations of unstressed and stressed vowels in initial, medial and final positions in the word. We looked at utterances in which the stressed vowel was /a/ (or /ʌ/), rather than /i/, so that, for the Italian speakers, stressed and unstressed vowels had the same phonetic quality, and, for English talkers, both were central vowels.

The results are presented in Table 5. Among Italian talkers, whereas stressed vowels were consistently longer than unstressed vowels (with a ratio between them of 1.75 on the average), unstressed syllables themselves showed essentially no variability across position in the word. (Final lengthening is sometimes found in Italian speech, however; see Vayra et al., 1984.) Although stressed vowels showed a greater range of variation, no consistent effects of word position were evident in stressed vowels either.

For their part, English talkers also showed consistently longer stressed than unstressed vowels, with an average ratio between them of 1.63. That the ratio is not larger for the stress-timed language is surprising (but see also den Os, 1985), although this may, in part, be ascribed to the apparently faster rate (and, therefore, overall shorter stressed vowels) among talkers of English. In contrast to Italian talkers, both English talkers had lengthened final syllables, whether stressed or unstressed.

TABLE 5
Durations of unstressed and stressed vowels in different positions in the word.

	Unstressed Vowel			Stressed Vowel		
	Initial	Medial	Final	Initial	Medial	Final
Italian						
CB	87	73	87	133	181	146
PMB	70	67	58	107	112	94
MC	109	111	109	174	204	206
English						
GD	55	55	79	86	96	120
CF	49	45	57	75	68	113

DISCUSSION

Our findings corroborate others in suggesting that several aspects of timing structure show a similar pattern in English. First, other studies show that English has a left-dominant foot structure (e.g., Selkirk, 1980) such that rules for constructing patterns of relative prominence in words assume a sub-word constituent structure consisting of a stressed syllable and following unstressed syllables. Second, patterns of shortening in English are asymmetrical with unstressed syllables following a stressed syllable associated with measured shortening of the latter. Preceding unstressed syllables are associated with only weak shortening or perhaps none at all. Third, coarticulatory influences of a stressed vowel on unstressed neighbors show a similar asymmetry whereby stressed syllables coarticulate more with following than with preceding unstressed syllables.

Fowler (1981) proposed that the English foot is, in part, an organization of stressed and unstressed syllables. In particular, stressed vowels serve as the articulatory foundation of the foot with following unstressed vowels superimposed on the trailing edge of a stressed vowel. This superimposition or ("coproduction") has two consequences. First the stressed vowel coarticulates substantially with following unstressed vowels and second, it is measured to shorten in feet that include unstressed syllables. Our present findings are compatible with that description.

Our present findings on English have one more implication of interest. In a recent paper, Jassem et al. (1984) compare two perspectives on English speaking rhythm. One is foot-based (Abercrombie, 1964) and the other is their own view that "total rhythmic units" consist not of feet, but of "narrow rhythmic units" and "anacruses." Anacruses are proclitic syllables that belong syntactically with the phonetic material in the following foot (Hill, Jassem, & Witten, 1978); narrow rhythmic units consist of any syllables that are not anacruses. Jassem et al. argue that their own system fits the durational data better because it distinguishes anacruses, which do not compress, with narrow rhythmic units, which do on average. However, Jassem et al. do not look at unstressed syllables in narrow rhythmic units separately from the unit's stressed syllable to determine where the compression occurs in narrow rhythmic units. Our findings are that unstressed syllables, in general, resist shortening, not just those that are anacruses.

Our findings on Italian are not consistent with those on English except in this last respect. Although there is evidence in the literature (Nespor & Vogel, 1979; but see

Bertinetto, 1981b; 1985) that a left dominant foot structure explains the pattern of relative prominence in Italian words, our study provides no evidence that the foot serves as an effective organizational constituent in Italian word production. Whereas Italian speakers do show coarticulatory influences of stressed on unstressed vowels, only one talker showed a reliable asymmetry consistent with a left-dominant foot. Likewise only one talker showed an appropriate shortening asymmetry. Finally, none of the talkers showed a significant relation between measures of coarticulation and shortening as they should if both measures were consequences of coproduction of stressed and unstressed syllables.

One implication of this outcome may be that compensatory shortening is not an important part of the prosody in Italian. A second is that coarticulation and shortening are not necessarily coupled in speech, as they would be, other things equal, were measured shortening a consequence of coarticulation.

Further cross-linguistic studies of metrical structure, coarticulation, and shortening across languages identified with the different timing typologies are required before firm conclusions can be drawn about the differences between Italian and English talkers. One conclusion that is possible given the findings to date is that the convergence of coarticulatory and shortening asymmetries in English is accidental—that is, coarticulation and perhaps shortening too do not reflect the foot structure of the language except accidently, and that measured shortening is not a by-product of coarticulation as coproduction. An alternative is that the convergence is not accidental, but is special to stress-timing and, in fact, underlies the impression that English (as contrasted with Italian) is stress timed. Other possibilities exist as well.

Of particular interest in an assessment of the relations among the variables, foot structure, coarticulation and shortening are studies of other languages identified as stress-timed, especiallly those that have right-dominant feet (see Hayes, 1981, for some examples and Wenk & Wioland, 1982, for a suggestion that French falls into this category). If coarticulation and shortening reflect a foot structure based on or reflected in temporal relations among syllables, then languages with right-dominant feet should show greater anticipatory than carryover coarticulation and greater backward than anticipatory shortening.

Study of other syllable-timed languages is necessary, too, in order to determine whether the individual differences and the generally disorderly relation between coarticulation and shortening that we observed among Italian speakers at (what would be) the foot level of description is general to languages having this timing type.

ACKNOWLEDGMENT

The Italian authors express their gratitude to Haskins Laboratories for their generous hospitality during the authors' stays there in 1980, 1981, and 1982 and to the Scuola Normale Superiore for supporting these stays. They are grateful also to G. Nencioni, A. Liberman, and M. Studdert-Kennedy for their encouragement and guidance throughout the project, to P. M. Bertinetto and G. Marotta for critical advice on the project, E. Gozzoli for his work on the waveform editing and display system at the Scuola Normale, the colleagues of the Centro di Studio per le Richerche di fonetica del CNR in Padova for the use of spectrographic facilities, to G. P. Teatini for many helpful discussions, and especially they thank B. Pedrazzi and A. L. Profili, former students at the Scuola di Logopedia of Ferrara University Medical Faculty, for helping with measurements and for their enthusiasm. This research is part of a larger project of contrastive studies on English and Italian rhythmical and intonational structures designed by all three authors.

In the present paper the part on English is mainly due to C. Fowler, the part on Italian to M. Vayra; C. Avesani contributed to the acoustical analysis of the data.

This research was supported in part by Grants HD-01994 and BNS-8111470 to Haskins Laboratories.

REFERENCES

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, & J. Trim (Eds.), *In honour of Daniel Jones*. London: Longman.

Bertinetto, P. M. (1981a). *Strutture prosodiche dell'italiano*. Firenze: Accademia della Crusca.

Bertinetto, P. M. (1981b). Ancora sull'italiano come lingua ad isocronia sillabica. In *Scritti linguistici in onore di Giovan Battista Pellegrini*. Pisa: Pacini.

Bertinetto, P. M. (1985). Recenti contributi alla prosodia dell'italiano. *Annali della Scuola Normale Superiore, II*, 581-644.

Classe, A. (1939). *The rhythm of English prose*. Oxford: Blackwell.

Cooper, W., & Eady, S. (1986). Metrical phonology in speech production. *Journal of Memory and Language, 25*, 369-384.

Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

Farnetani, E., & Kori, S. (1982). Lexical stress in spoken sentences: A study on duration and vowel formant patterns. *Quaderni del Centro di Studio per la Recerche di Fonetica, 1*. Padova: Progetto.

Farnetani, E., & Kori, S. (1983). Interaction of syntactic structure and rhythmical constraints on the realization of word prosody. *Quaderni del Centro di Studio per le Ricerche di Fonetica, 2*. Padova: Progetto.

Farnetani, E., & Kori, S. (1984). Effects of syllable and word structure on segmental durations in spoken Italian. *Quaderni del Centro di Studio per le Ricerche di Fonetica, 3*. Padova: Progetto.

Fowler, C. (1981). A relation between coarticulation and compensatory shortening. *Phonetica, 38*, 35-40.

Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General, 112*, 386-412.

Hayes, B. (1981). *A metrical theory of stress rules*. Bloomington: Indiana University Linguistic Club.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15*, 33-74.

Hill, D.R., Jassem, W., & Witten, I. (1978). A statistical approach to the problem of isochrony in spoken British English. *Man-Machine Systems Laboratory Report*, University of Calgary.

Jassem, W., Hill. D., & Witten, I. (1984). Isochrony in English speech: Its statistical validity and linguistic relevance. In D. Gibbon & H. Richter (Eds.), *Intonation, accent and rhythm*. Berlin: W. De Gruyter.

Kohler, K. (1983). Prosodic boundary signals in German. *Phonetica, 40*, 89-134.

Lepschy, G. C. (1966). I suoni dell'italiano: Alcuni studi recenti. *L'Italia dialettale, 29*, 49-69.

Lepschy, G. C. (1975). L'insegnamento della pronuncia italiana. *Studi italiani di linguistica teorica ed applicata, 4*, 201-209.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 11*, 51-62.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm, 21*, 1-59.

Marotta, G. (1984). *Aspetti della struttura ritmico-temporale in italiano*. E.T.S., Pisa.

Nespor, M., & Vogel, I. (1979). Clash avoidance in Italian. *Linguistic Inquiry, 11*, 467-482.

Nooteboom, S. G. (1973). Perceptual reality of some prosodic durations. *Journal of Phonetics, 1*, 24-45.

den Os, E. (1985). Vowel reduction in Italian and Dutch. *Progress Report of the Institute of Phonetics, University of Utrecht (PRIPU), 10*, 3-12.

Pike, K. (1945). *Intonation of American English*. Ann Arbor: University of Michigan Press.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry, 11*, 563-605.

Shen, Y., & Peterson G. G. (1962). Isochronism in English. *Studies in Linguistics, 9* (University of Buffalo Occasional Papers).

Vayra, M., Avesani, C., & Fowler, C. (1984). Patterns of temporal compression in spoken Italian. In M. Van den Broecke & A. Cohen (Ed.), *Proceedings of the Tenth International Congress of Phonetic Sciences*. Dordrecht, Holland: Foris Publications.

Wenk, B., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10*, 193-216.

FOOTNOTES

Studi di Grammatica Italiana, XIII, 243-269, in press.

†Scuola Normale Superiore, Pisa, Italy.

††Also Dartmouth College.

[1]Previous studies have looked at Standard Italian speakers of other dialectal areas; for example: Venetia (Farnetani & Kori, 1982) Lombardy (Farnetani & Kori, 1984); Tuscany (Farnetani & Kori, 1984; Marotta, 1984; den Os, 1985).

[2]"Dall' a me" is a phrase with an elided vowel "o" after the first syllable.

[3]Pseudowords were different for English and Italian speakers because the data for English talkers were not collected originally with this comparison in mind.

[4]Both talkers produced this word with four syllables.

Awareness of Phonological Segments and Reading Ability in Italian Children*

Giuseppe Cossu,** Donald Shankweiler,† Isabelle Y. Liberman,† Giuseppe Tola,†† and Leonard Katz†

The early evidence pertaining to the development of phonological segmentation abilities and their relation to reading was collected with English-speaking subjects. Although data from other languages have been obtained, explicit cross-language comparisons have not been made. It was considered that since languages vary in their phonological structures, they may also vary in the demands they make on the beginning reader. The present study compared the segmentation abilities of Italian children with those of English-speaking (American) children using the same methods of assessment and the same subject selection criteria. At the preschool level, though the Italian children manifested a higher level of performance overall, their pattern of performance paralleled that obtained earlier with American children. In both, syllable segmentation ability was stronger than phoneme segmentation. After school entrance, this pattern remained unchanged in American children, but was reversed in Italian beginning readers.In both language groups, however, phonemic segmentation ability distinguished children of different levels of reading skill. The discrepancies between the language groups were seen as reflecting phonologic and orthographic differences between the languages.

Those who would become proficient readers of a language that is written with an alphabet face a common problem: they must understand that the written letters represent segments of words. For this reason, mastery of an alphabetic system requires a metalinguistic capability that is quite unnecessary for acquisition of the spoken language, namely, some degree of metalinguistic awareness that words have those sublexical segments, the phonemes (Liberman, 1971, 1973). We know that very young children may not have that capability. In fact, awareness of phonemic segmentation, for many English-speaking children, at least, is delayed until age six or beyond (Liberman, Shankweiler, Fischer, & Carter, 1974). Experimental evidence from a variety of sources suggests that awareness of phonemic constituents of a word is highly correlated with reading achievement (Stanovich, 1982; Wagner & Torgesen, 1987). In fact, poor beginning readers and illiterate adults both tend to find the phonemic structure of spoken words quite opaque (Liberman, Rubin, Duqués, & Carlisle, 1985).

Although a large part of the evidence pertaining to the development of segmentation abilities and their relation to reading has been collected with English-speaking subjects (see Bradley & Bryant, 1983; Fox & Routh, 1976; Liberman et al., 1974; Treiman & Baron, 1981), data from speakers of other languages have begun to be collected. In addition to the American and British studies of preschoolers and school children in the elementary grades, studies of Yugoslavian speakers of Serbo-

Croatian (Ognjenović, Lukatela, Feldman, & Turvey, 1983), Swedish (Olofsson, 1985), French-speaking (Alegria, Pignot, & Morais, 1982) and Spanish-speaking children (de Manrique & Gramigna, 1984) have also been carried out with roughly similar results. The failure of illiterate adults to perform phoneme segmentation was first demonstrated with speakers of Portuguese (Morais, Cary, Alegria, & Bertelson, 1979). Similar findings have since been obtained in the U.S. with English-speaking semi-literate adults (Liberman et al., 1985), and in China with readers of Chinese logograms who were unacquainted with the alphabet (Read, Zhang, Nie, & Ding, 1984).

Thus, studies of subjects from different language backgrounds have provided considerable support for the possibility of a significant relationship between phoneme segmentation and the mastery of the alphabetic principle. However, many questions that might be answered by cross-language studies have not yet been systematically explored. We know that languages vary widely in the complexity of their phonological structure. They may vary, for example, in their number of distinguishable vowels, in the incidence of morphophonemic alternation, and in the diversity of their syllable types. Moreover, alphabetically written languages differ not only in the complexity of their phonologic structure, but also in the ways in which the orthography chooses to transcribe that structure (Klima, 1972; Liberman, Liberman, Mattingly, & Shankweiler, 1980). It would not be unreasonable, therefore, to ask whether the ability to analyze words into their component segments might be harder to acquire in some languages than in others.

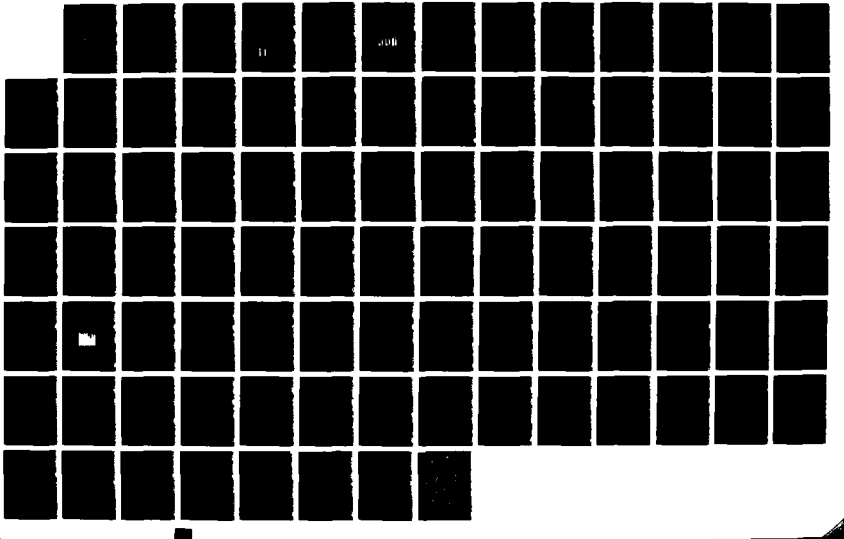
Further, it could be asked whether variations in phonological structure that affect the ease or difficulty of becoming aware of critical sublexical units do in fact also affect the ease or difficulty of learning to read (Liberman et al., 1980). Cross-language comparisons that take into account the nature of both the language and the orthography can thus be of particular importance in sorting out the root causes of reading problems.

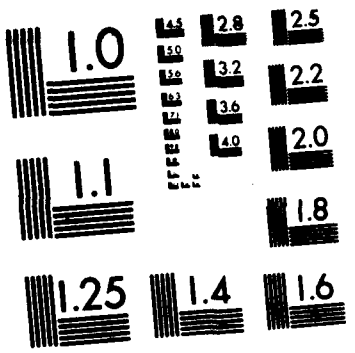
Although the problems of acquiring literacy are increasingly being studied in different language communities, few actual cross-language comparisons have been undertaken of the development of phoneme segmentation and its relation to reading. In view of the evidence to date concerning the critical role of the phoneme in learning to read and write alphabetically, we considered it an urgent priority to conduct a parallel investigation in another alphabetic language with different structural features than English. The present investigation is an attempt to replicate as closely as possible in Italian previous studies undertaken with English-speaking children.[1] Only by such a comparative study can we hope to distinguish biological maturational factors relevant to acquiring literacy skills from task factors that reflect specific features of a particular language and its writing system. Thus, eventually, we may learn whether it is universally true, as we suspect, that the phoneme is a particularly difficult unit for young children to abstract. Further, we may learn whether the ability to abstract the phoneme relates positively to reading success in any alphabetic system irrespective of differences in language structures and their means of representation in the orthography.

Italian is a useful candidate for comparison with English because it differs markedly in certain aspects of phonological structure. Let us consider several ways in which it differs. The first is in vowel structure. Spoken Italian can be fairly said to have only five vowels; spoken English has a dozen or more (depending on the criteria for analysis). Regardless of the context in which they occur, each of the five vowels in the alphabet has only one rendition in Italian speech as contrasted with the several that can be found in English. We would not expect this contrast to materially affect the relative difficulties of phoneme and syllable awareness in the two language communities before reading instruction is begun. For preschoolers, the coartic-

AD-A192 001 STATUS REPORT ON SPEECH RESEARCH(U) HASKINS LABS INC NEW HAVEN CT M STUDDERT-KENNEDY SEP 87 SR-91(1987) PHS-HD-01994 2/2

UNCLASSIFIED F/G 5/7 NL





ulation of the sublexical phonemes in normal speech should still make phoneme awareness relatively more difficult than syllable awareness, whatever the language. Once children are exposed to the orthography, however, it is possible that linguistic awareness of both kinds might be better among the Italian children, because of the nature of the Italian vowels and their relation to the orthography.

Italian and English vary in other characteristics besides vowel structure that may affect phonological awareness differentially after reading instruction has been initiated. Italian has a relatively shallow phonology with, for example, relatively little morphophonological alternation as compared with English (e.g., telegraph, telegraphy). In addition, though Italian has a mixed stock of syllables types, it has fewer than half as many different types as English (Carlson, Elenius, Granström, & Hunicutt, 1985). Moreover, unlike English, which has a predominantly closed syllable structure (e.g., CVC,CVCC,CCVC, etc.), Italian's most frequent syllable form by far is the open syllable (e.g., CVCVCV,CVCV, etc.) with relatively few different variations (Carlson, et al., 1985). It has been suggested (Liberman & Shankweiler, 1979) that because the syllable is the basic unit of articulation, it has a perceptual saliency that allows it to be easily extracted from the speech stream. It is possible that given the particular characteristics of its syllable structure, Italian might lend itself even more readily than English to sublexical awareness at the syllabic level. The simpler syllable structure, the smaller number of distinctively different vowels, and the greater consistency of the alphabetic representation in Italian might also be expected to give Italian an advantage in early reading acquisition that would in turn be reflected in greater phoneme awareness as well. The reciprocal effect of reading acquisition on phoneme awareness has been frequently proposed (see, for example, Vellutino, 1979).

The present study consists of two experiments: Experiment 1 addresses questions relating to the development of metalinguistic awareness of sublexical segments. Experiment 1 allows us to tease apart the maturational component in the development of metalinguistic awareness of sublexical segments from the possible contribution of differences in language structure. Accordingly, it addresses the following questions: 1) Is the level of success in abstracting either syllabic or phonemic segments consistently higher for one language than the other? 2) Does ability to abstract either type of segment vary with age in the same manner in Italian children as in their English-speaking counterparts? Experiment 2 is concerned with the relation between metalinguistic awareness and reading acquisition. It is directed at children of varying reading skill in the first two elementary school grades. It asks first whether the ability to single out syllable and phoneme segments is related to level of reading achievement in Italian children, as was found for English-speaking children, and second, whether a relatively straightforward orthography like Italian facilitates the acquisition of awareness of phonological segments more than a relatively complex orthography like English.

EXPERIMENT 1: Development of Segmentation Ability

Subjects

The subjects included two samples of children from a largely middle-class school in the Sardinian town of Sassari (Italy): 60 preschool children and 160 school children from the first and second elementary classes. Children with known auditory, visual, language, or motor deficiencies were excluded from the sample, as were those with clinical histories indicating brain damage.

The preschool sample included 60 children from the regular second- and third-year preschool classes (roughly comparable to American nursery and kindergarten levels). They were then divided by class membership into two groups of 30 children,

each of which contained equal numbers of boys and girls. The mean age of the younger group, Group A, who were members of the second-year preschool class, was 52.9 months, range 48 to 59. The older group, Group B, which included the third-year preschool class, was roughly 16 months older—mean age 68.8 months, range 62 to 72. Groups A and B were each further divided into subgroups, those given syllable (Syl) or phoneme (Pho) segmentation tasks. The mean ages of the subgroups were as follows: Group A-Syl, 51.8; Group A-Pho, 54; Group B-Syl, 69; Group B-Pho, 68.6.

The level of intelligence of the preschool subjects was measured by the Goodenough Draw-a-Person Test (DAP). This was the measure of intelligence used in the comparison study (Liberman et al., 1974). The mean IQs on the DAP were as follows: Group A-Syl, 98.6; Group A-Pho, 105.8; Group B-Syl, 102.1; Group B-Pho, 101.2. Across preschool classes, the mean IQ was 102.2 for Group A, the younger class, and 101.6 for Group B, the older class.

The elementary school sample included two groups of 80 children each (half boys, half girls) attending, respectively, the first grade (mean CA 84.3) and the second grade (mean CA 96.8). The mean ages across segmentation tasks and grade were practically identical for the first grade: 84.2 and 84.3, respectively, for the Pho and Syl tests, and for the second grade, 96.8 for each task group.

The level of intelligence of the elementary school children was assessed by the Verbal Scale of the Wechsler Intelligence Scale for Children (WISC). When computed across tasks, the mean IQ was 107.2 for the syllable group, 107.4 for the phoneme group. When computed across grade levels, the IQs for first and second grades were 109.2 and 105.4, respectively.

Procedure

The procedures were modeled as closely as feasible after the procedure of Liberman et al. (1974). Under the guise of a "tapping game," the child was required to repeat a word spoken by the examiner and then to indicate by tapping a small wooden dowel on the table, the number of (from two to four)[2] segments (phonemes for group Pho and syllables for group Syl) in the stimulus items. The test items were spoken by the examiner (and repeated by the child) in a natural manner. Each child received only one of the two types of tasks. Instructions were the same for all the subjects in both the preschool and elementary school groups.

Procedures for the two experimental groups followed an identical format, differing only in the test items used for the two tasks. Four sets of training trials containing three items each were given. During training, each set of three items was first demonstrated in an order of increasing complexity (from two to four segments). When the child was able to repeat and tap each item in the triad set correctly as demonstrated in the initial order of presentation, the items of the triad were presented individually in scrambled order without prior demonstration, and the child's tapping was corrected as needed. The test trials, which followed the four sets of training trials, consisted of 45 randomly assorted individual items of two, three, or four segments that were presented without prior demonstration and were corrected by the examiner, as needed, immediately after the child's response. Testing was continued through all 45 items[3] (see Appendix 2 for the stimulus items). Each child was tested individually by the same examiner in a single session near the end of the school year.

Materials

The stimulus materials, including training and test trials, were the same for both preschool and elementary grade children. For the 15 two- and three-syllable words as well as for the three- and four-phoneme words, the stress was always on the first syllable. Among the four-syllable words, five had the stress on the second syllable

and ten on the third. This uneven distribution reflects the frequency of occurrence in Italian.

Results

The question of whether there are overall differences between the language groups in level of performance on metalinguistic tasks can be assessed by the same scoring procedure employed with English-speaking children by Liberman et al. (1974). In that procedure, the comparison was made in terms of the percentages of four-, five-, and six-year olds who reached a criterion of six correct trials without demonstration in syllable and phoneme segmentation tasks. When we carry out this same procedure, we find that the pattern of performance in the two language groups is similar but the success rate is quite different. As can be seen in Table 1, a higher proportion of the Italian children at each age level succeeded on each task. Even from the nursery level, the markedly greater percentage of Italian children reaching criterion is apparent. Syllable analysis was relatively easy as compared to the phoneme even for the English-speaking children, but the advantage of syllable over phoneme becomes even more striking with the Italians, especially at the preschool levels.[4]

TABLE 1

Percentage of Italian-speaking and English-speaking* Children Reaching a Criterion of Six Successive Correct Responses Without Demonstration.

	TASK			
Age	Phoneme		Syllable	
	Italian	*English*	*Italian*	*English*
Nursery	13	0	67	46
Kindergarten	27	17	80	48
First Grade	97	70	100	90

*Data from Liberman et al. (1974).

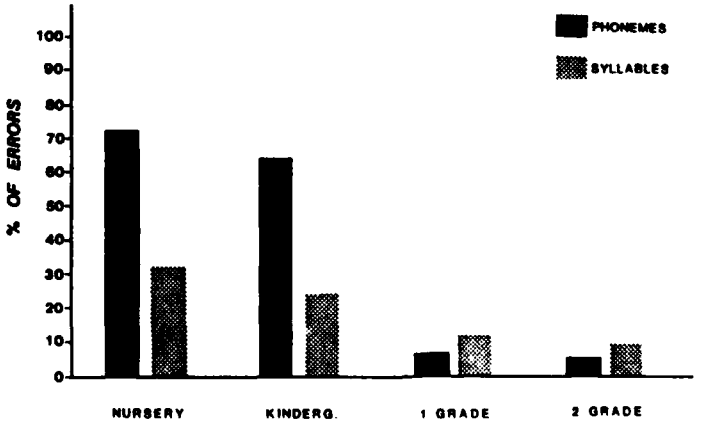


Figure 1. Percent errors in segmentation on the phoneme test and the syllable test in children at four age levels.

In assessing further the question of whether the ability to abstract either type of segment varies with age in the same manner in Italian children as in their English-

speaking counterparts, we tabulated the overall number of errors made by each child on the 45 items. The results of this tabulation are displayed in Figure 1. Considering the findings for all 220 children, it is apparent that there is, as would be expected, a marked improvement in performance level with increasing age. The errors of phoneme segmentation range from a high of 72.6% in four-year olds to 65.6% in kindergarteners, followed by a striking drop to 6.9% in the first elementary school grade and then a slight decrease of 5.8% in the second grade. Performance on syllable segmentation shows a similar falling trend across grade levels, but initial performance is much better and progress is more gradual (nearly linear) from grade to grade, decreasing from 35.1% in four-year-olds to 9.4% in the seven-year-olds (second graders).

The total error scores were subjected to a two-way analysis of variance. The effect of grade, $F(3,212) = 168.02$, $p < .0001$, task, $F(1,212) = 17.84$, $p < .0001$, and the interaction of task and grade, $F(3,212) = 45.86$, $p < .0001$, are all highly significant.

The significant effects of grade and task were anticipated, but the interaction is a new outcome. As may be seen from Figure 1, the interaction may result in part from an apparent reversal in the relative difficulty of the two tasks for the preschool and the school-age groups. As we noted earlier, the preschoolers found the Pho task much more difficult than the Syl task, just as their American counterparts did. In contrast to the preschoolers, the first and second grade elementary school children made relatively few errors on either task, but at both the first and second grade levels, there were more errors on the Syl task than on the Pho task. Comparison of the differences by post-hoc t-tests shows them to be significant (first grade: $t(212) = 3.70$, $p < .0$; second grade: $t(212) = 2.99$, $p < .005$), an outcome to which we will return in the Discussion.

EXPERIMENT 2: Segmentation Abilities in Readers of Varying Skill

For the purposes of this experiment the 80 elementary school subjects from grades one and two were divided into groups of good, average, and poor readers by grade. A reading test consisting of 60 bisyllabic words derived from word lists in first and second grade reading texts (Carlino-Bandinelli, 1984) was used to assess reading achievement. For each grade, the 30 best and 30 poorest achievers were selected, leaving the remaining 20 as average readers. One half of each group was allocated to the Syl task and one half to the Pho task. Mean ages and IQs for these subgroups are given in Table 2.

TABLE 2

Mean Age and IQ by Grade Level, Task, and Reading Achievement

		FIRST GRADE (n=80)		SECOND GRADE (n=80)	
		Mean Age	IQ	Mean Age	IQ
GOOD READERS (n=30)	Pho	83.1 [79-92]	115.0 [89-135]	96.3 [91-101]	105.8 [80-128]
	Syl	84.4 [79-89]	106.9 [89-131]	98.2 [93-104]	105.8 [81-130]
AVERAGE READERS (n=20)	Pho	85.2 [75-90]	118.3 [81-138]	94.9 [89-103]	108.2 [80-142]
	Syl	84.8 [81-90]	106.3 [80-130]	95.0 [87-102]	115.0 [97-142]
POOR READERS (n=30)	Pho	84.6 [78-91]	103.0 [90-115]	98.5 [93-104]	198.1 [84-113]
	Syl	84.0 [77-92]	107.6 [90-140]	96.5 [90-102]	103.8 [84-131]

Results

We now turn to further analysis of the data for first and second elementary grade children whose overall performance was depicted in the right hand portion of Figure 1. Figure 2 displays their performance on the Syl and Pho tests separately for subgroups of good, average, and poor readers.

The differences between the subgroups were evaluated by a two-way analysis of variance with task (Syl and Pho tests) and level of reading achievement as factors. Both factors are significant: task, $F(1,148) = 8.63$, $p < .004$; reading level, $F(2,148) = 4.27$, $p < .02$. There was no interaction between them.

In spite of the absence of a significant interaction, two striking characteristics of Figure 2 prompted a more detailed assessment of differences in the performance of the reading achievement subgroups at grade 1 and grade 2. First, inspection of the figure suggests that the average readers improved in phoneme segmentation from Grade 1 to Grade 2, although the Good and Poor readers did not. This suggestion was supported statistically by t-tests: for Average readers, $t(148) = 3.40$, $p < .001$ for the Pho task, while the results for Good and Poor readers were nonsignificant. For the Syl task, the Poor readers improved significantly from Grade 1 to Grade 2, $t(148) = 4.53$, $p < .001$, while for the other two reading achievement groups, the differences were nonsignificant. Thus we find that some changes in performance of each task occur between first and second grade. But since the changes are specific to reading level, they are more powerfully detected when we examine the reading-achievement subgroups separately.

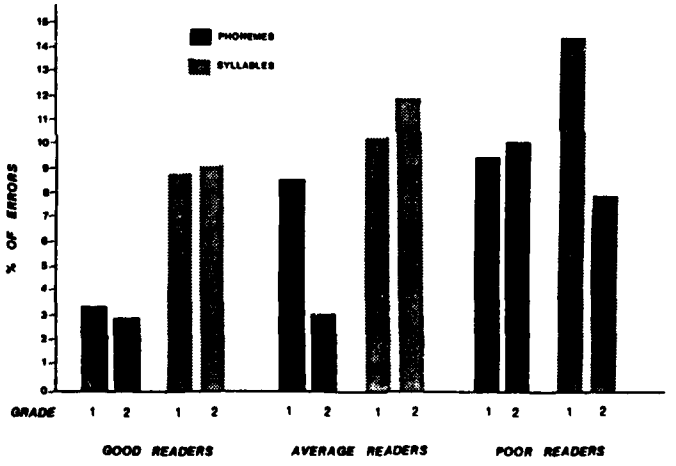


Figure 2. Percent errors in segmentation on the phoneme test and the syllable test in first- and second-grade children grouped by reading ability.

Discussion

Our concern in this investigation was to begin to sort out the effects of specific language characteristics on the development of metalinguistic awareness and the early stages of reading acquisition. To this end, and to eliminate confounding by procedural differences, we elected to attempt a direct replication with Italian children of an earlier study of American children (Liberman et al., 1974). Accordingly, both in subject selection and in experimental design, the two studies are as similar as was practicable. But even so, in our attempt to make the Italian study match its American counterpart, we were thwarted to some degree by the very nature of the language differences themselves. For example, because Italian contains very few monosyllabic words, it was not possible to construct a phoneme judgment task

with monosyllables, as had been done in the English-language study. Similarly, since high-frequency Italian words tend to contain greater numbers of syllables than their counterparts in English (Carlson et al., 1985), the syllable test, which in the Englishversion contains one-, two, and three-syllable words, contains two-, three-, and four-syllable words in the Italian version. Though these differences might be expected to make the stimuli more difficult than in the American study, they apparently did not have a deleterious effect on performance, as we will see later. One further difference, but one that does not affect the comparisons we are making, is that the Italian study was expanded in scope to include second graders in addition to children in preschool and first grade classes.

In discussing the findings we will consider first the question of the development of metalinguistic awareness that was pursued in Experiment 1 and then the relation to reading achievement, which was investigated in Experiment 2.

Development of Awareness of Phonological Segments

The present study demonstrated evidence of cross-language similarities as well as differences concerning the development of metalinguistic skills in phonological segmentation. Let us first note the similarities. At the preschool level, the present study confirmed an improvement with age from nursery school to kindergarten, a progression that is consistent with a view that the early development of segmental analysis ability is under maturational control. It also confirmed that the greater difficulty of abstracting phonemic units than syllable units applies across languages that differ in their phonological structure. Up to the first grade level, the findings on Italian children in Experiment 1 yield a pattern of results on both the Syl and Pho tests that is essentially consistent with the results on American children. At both preschool ages, the Italian children, like their American counterparts, had relatively little success in identifying phonemes and were more successful in identifying syllables.

To account for the earlier accessibility of the syllable in both language groups, we appeal to the suggestion by Liberman (1971), that the syllable may be easier to identify, whatever the language, because it is a temporally discrete phonetic unit, whether speech is considered from an articulatory or an acoustic point of view. Unless they already happen to be syllables, most phonemes, in contrast, have no independent existence, being always assimilated by coarticulation. On these phonetic grounds it makes sense that most phonemic segments would be harder to bring to consciousness.

Even after reaching school age, the Italian children of the present study show a basic similarity to their American counterparts. In both language communities, a sharp improvement in performance follows instruction in reading. In both, there was a marked decline in errors in both segmentation tasks after a short period of reading instruction. It is therefore judged that the present findings give further credence to the suggestion (see Read et al., 1984) that exposure to an alphabetic orthography has a positive effect on metalinguistic awareness of phonemic segmentation.

So much for the similarities. The present study also brought to light several differences from its English parallel, which may be ascribed in part to structural differences between the languages. At the preschool level, we found a quantitative difference in the degree of accuracy in the two language groups—the Italians made fewer errors on both tasks. It is notable also that the Italian children performed more accurately despite having to deal with items containing greater numbers of syllables. We may speculate as to the reason for this difference in performance. We have suggested that because of the simpler open-syllable structure, the small number of syllable types and vowel distinctions in Italian, segmental analysis into syllables would be easier than in a language like English with its closed-syllable structure and

more numerous syllable types and vowel distinctions. Similar results should therefore be attained in other languages that are similar in structure to Italian, like Spanish—an expectation that is borne out in the research of de Manrique and Gramigna (1984).

The level of performance of both groups improves markedly with first grade attendance. But there is a difference. By first grade, the Italian children are at ceiling on both tasks, but the Americans are not. The difference is most marked on the phoneme task where only about 3% of the Italian children failed to reach criterion in contrast to their English-speaking counterparts, 30% of whom still fail the phoneme test at this age.

Several factors may account for the relative superiority of the Italian children's performance. These relate both to language structure and the orthography. As we have said, the uniform, open-syllable structure and the smaller number of different vowel sounds in the language must surely make the basic analytic task easier. Once reading instruction is initiated, the closer correspondence between letters and phonemes in the Italian orthography should further facilitate the child's development of sensitivity to sublexical structure. There is more often than not a closer, one-to-one correspondence between letters and phonemes in Italian than is common in English. In consequence, it is fair to say that if one knows how to spell an Italian word, its phonemic makeup is almost automatically revealed.

In Figure 1, we note a further disparity with the results of Liberman et al. (1974): the errors on the Pho test for Italian first graders actually drop below the level of error on the Syl test, whereas for the American first graders, the Pho error rate, though dropping, remained higher than that for the syllables. (Although the error rate continues to diminish somewhat in the Italian second grade, the same reversal is evident.)

Given that for children younger than elementary school age, phonemic segmentation was still so much more difficult than segmentation by syllable, what accounts for the reversal of difficulty for the Italian first (and second) graders? Our study does not provide a final answer to that question. Since there are so few errors of either phoneme or syllable segmentation in the Italian data, we probably need not be too concerned about the difference. However, we may speculate that given its fairly regular orthography, the small number of vowels, and the open syllable structure, reading and spelling practice in Italian may actually provide daily phonemic training for the learner, since each phoneme would be readily identified with a particular letter (and vice versa). If this line of reasoning is correct, skill in phonemic segmentation could increase at a faster rate than syllabic segmentation skill, thus accounting for the apparently paradoxical situation of a reversed pattern between Pho and Syl skills. The difference is particularly evident in the performance of the better readers, who presumably have made best use of the information provided.

Awareness of Phonological Segments and Reading Ability

We now turn to consider the analysis of the effects of differences in reading attainment on segmentation skills at each grade level. In the Italian study, unlike the parent American study, we have data for both first and second school years, and for average readers as well as the extremes. As may be seen in Figure 2, the relationship between level of performance on the two segmentation tasks is different for good, average, and poor readers: the average group, not the good readers or the poor readers, showed the greatest improvement in phoneme analysis between first and second grade. Conceivably, this is the group that benefited most from continued instruction. The good readers may not have benefited so much because their performance was already near the asymptote. The poor readers may not have benefited because they still had unresolved problems with the orthography.

On the Syl test, on the other hand, it is the poor readers who appear to gain with experience. Their relatively high error rates in first grade are consistent with initial confusion about how this unit relates to the orthography. It is of interest in this regard that the first grade poor readers have a striking tendency to overestimate the number of syllables in a word. That is, an error analysis revealed that when poor readers made a mistake, their errors were overestimates three times more often than underestimates. This is consistent with the possibility that they are overestimating the number of syllables because they are sometimes confusing syllable structure with phonemic structure.

The question arises whether IQ can account for the differences we found on segmentation performance. Although there are differences among the age groups and reading ability groups on IQ, there is no consistent relationship between IQ and number of errors either on the Syl or the Pho test. For example, for Grade 1, the number of Pho errors is least for Good Readers (3.5%) and greater for Average Readers (8.5%) and Poor Readers (9.5%), but mean IQ for these groups does not decrease consistently with error rate. Instead, the Good Readers have a lower IQ (115.0) than the Average Readers (118.3), with the lowest IQ appearing for the Poor Reader group (103.0). Similarly, Pho errors for the second graders are nonmonotonically related to IQ. Finally, in the one case where Syl errors do show an apparent relation with IQ—among second graders—the direction of the relationship is counterintuitive: higher IQ is associated with the greater number of syllable errors. Thus, it seems implausible that IQ is an important factor.

In summary, the present study demonstrated evidence of cross-linguistic similarities as well as differences concerning the development of metalinguistic skills in phonological segmentation and their relationship to reading ability. At the preschool level, the present study confirmed that the greater difficulty of abstracting phonemic units than syllable units applies even across languages that differ in their phonological structure. In both language communities, a sharp improvement in performance also follows instruction in reading. In both, there was a marked decline in errors in both segmentation tasks after a short period of reading instruction. Concerning the relationship between level of reading skill and segmentation ability, Italian poor readers, like their American counterparts, are differentiated from the good readers by the phoneme test.

The present study, however, also brought to light several differences from its English parallel, which may be ascribed in part to structural differences between the languages. At the pre-school level, we found a quantitative difference in the degree of accuracy between the two language groups—the Italians made fewer errors on both the syllable and phonemic tasks. In regard to the syllable task, the open-syllable structure of the Italian language was thought to play a facilitative role. In regard to the phoneme task, that factor in combination with the smaller number of vowels and relatively more shallow orthography may be relevant.[5]

The most striking disparity between the two language groups, however, appeared after school entrance. The decline in errors of both types was considerably more marked in the Italian sample. Moreover, with the decline there appeared in the Italian children a pattern of performance opposite to that displayed by both the Italian preschool group and its English counterpart. For the Italian first graders, syllable analysis now produced more errors than phonemic analysis. A similar reversal is evident in the Italian second grade, where there are no comparable data in the English sample. We have speculated that differences in language and orthography are again at work. Of course, there is always the possibility that differences in teaching method should be held responsible to some degree for the differences found here. This is an unlikely explanation, however, since instruction in the Sardinian public schools, like the American, is an eclectic mixture of different approaches.

Future research in other language communities and careful control of teaching methods will be needed to explore these possibilities further.

ACKNOWLEDGMENT

The research and the preparation of the manuscript were supported in part by a Program Project Grant (HD-01994) to Haskins Laboratories from the National Institute of Child Health and Human Development. We are grateful for the helpful criticisms of two colleagues, Stephen Crain and Anne Fowler, and we would like to acknowledge also our indebtedness to an anonymous reviewer for penetrating comments that enabled us to clarify our presentation of this research.

REFERENCES

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory & Cognition, 10,* 451-456.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature, 301,* 419-421.

Carlino-Bandinelli, A. (1984) *Cominciamo 1° 2°,* Istituto Geographico De Agostini, Novara.

Carlson, R., Elenius, K., Granström, C., & Hunnicutt, S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. *Quarterly Report* (KTH Speech Transmission Laboratory), *1,* 63-94.

de Manrique, A. M. B., & Gramigna, S. (1984) La segmentacion fonologica y silabica en ninos de prescolar y primero grado. [Phonologic and syllabic segmentation in preschool and first grade children]. *Lectura y Vida, 5,* 4-13.

Fox, B., & Routh, D. K. (1975). Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistic Research, 4,* 331-342.

Fox, B., & Routh, D. K. (1976). Phonemic analysis and synthesis as word attack skills. *Journal of Educational Psychology, 68,* 70-74.

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye.* Cambridge, MA: MIT Press.

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society, 21,* 71-87.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23,* 65-77.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Liberman, I. Y., Rubin, H., Duqués, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergarteners and adult poor spellers. In D. B. Gray & J. F. Kavanagh, (Eds.), *Biobehavioral measures of dyslexia.* Parkton, MD: York Press.

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. Resnick & P. Weaver (Eds.), *Theory and practice of early reading* (Vol. 2, pp. 109-132). Hillsdale, NJ: Lawrence Erlbaum Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Lindgren, S. D., De Renzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development, 56,* 1404-1417.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition, 7,* 323-331.

Ognjenović, V., Lukatela, G., Feldman, L. B., & Turvey, M. T. (1983). Misreadings by beginning readers of Serbo-Croatian. *Quarterly Journal of Experimental Psychology, 35A,* 97-109.

Olofsson, A. (1985). *Phonemic awareness and learning to read. A longitudinal and quasi-experimental study.* Unpublished doctoral dissertation, Umea University (Sweden).

Read, C., Zhang, Y., Nie, H., & Ding, B. (1984). *The ability to manipulate speech sounds depends on knowing alphabetic spelling.* Paper presented at the 23rd International Congress of Psychology, Acapulco, Mexico.

Stanovich, K. E. (1982). Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32-71.

Treiman, R. A., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3). New York: Academic Press.

Vellutino, F. R. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press.

Wagner, R. R., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

FOOTNOTES

**Applied Psycholinguistics,* in press.

**University of Parma, Institute of Child Neurology and Psychiatry.

†Also University of Connecticut.

††University of Sassari, Institute of Child Neurology and Psychiatry.

[1]An earlier Italian-English comparative study by Lindgren, De Renzi, and Richman (1985) was an epidemiological investigation that had quite different objectives from the present research.

[2]The segment-counting tests of Liberman et al. (1974) contained items with one to three segments. The scarcity of words of one phoneme or one syllable in Italian necessitated the change. Language differences also prevented forming the triads for the training trials in the Italian study in the same systematic fashion with which they were formed in the English-language materials.

[3]In the Liberman et al. (1974) study, testing for each child was continued through the entire list or until the child reached the criterion of tapping six consecutive items correctly without demonstration. In the present study, all children were given the entire list but data are available for direct comparison of mean number of errors to passing (or failing) the criterion and of the mean number of children reaching criterion in both tasks.

[4]The same pattern of differences was noted across the Italian and American studies when the data were tabulated in terms of trials to reach criterion, and in terms of mean errors to passing and failing the criterion.

[5]Because of its many phonological similarities to Italian, the Spanish language would be a place to seek confirmation of these hypotheses. As it happens, the Spanish study by de Manrique and Gramigna (1984), to which we referred earlier, might have been an appropriate candidate for such a comparison, particularly since, in regard to its stimulus materials at least, it was closer to the original Liberman et al. (1974) study than our Italian replication. Unfortunately, several procedural differences in the Spanish investigation prevent direct comparison with either of the other studies; these differences include absence of a four-year-old group, sampling at the beginning rather than the end of the school year, and special syllable segmentation training provided for the kindergarteners. It can be said, nonetheless, that the findings are in general agreement as to the relative difficulty of the phoneme over the syllable for the preschool ages.

APPENDIX 1

Training Trials (Number of segments in parentheses)

PHONEMES			SYLLABLES		
(2)	(3)	(4)	(2)	(3)	(4)
DO	ALA	CENA	RETE	LIMONE	SERENATA
SO	APE	URTO	CASA	PORTONE	ELEFANTE
NE	UVA	ALTO	LUNA	ALBERO	TELEFONO
LI	ZIO	NOCE	ASTA	CAVOLO	RITORNELLO

APPENDIX 2

Test Trials (Number of segments in parentheses)

PHONEMES

LUCE (4)
MI (2)
VOLO (4)
ORA (3)
SU (2)
LE (2)
PER (3)
MA (2)
BAR (3)
FUMO (4)
NOI (3)
SI (2)
DUE (3)
FARO (4)
AL (2)
ARTE (4)
CON (3)
BA (2)
FA (2)
UNO (3)
STOP (4)
SIR (3)
TU (2)
DI (2)
ERBA (4)
DAL (3)
SE (2)
IERI (4)
VASO (4)
SEI (3)
TAC (3)
PIPA (4)
MAI (3)
SALE (4)
ARCO (4)
CI (2)
TRA (3)
NO (2)
TUBO (4)
UNTO (4)
SUO (3)
BLU (3)
VA (2)
OTRE (4)
IN (2)
PAPAVERO (4)

SYLLABLES

TERMOMETRO (4)
BARBA (2)
AMICO (3)
UOVA (3)
TEMPORALE (4)
MURO (2)
SARTO (2)
FUTURO (3)
BIRRA (2)
LUMACA (3)
INSALATA (4)
LAMPADINA (4)
ENTRARE (3)
GATTO (2)
CULLA (2)
PRETE (2)
TOPOLINO (4)
RAPINARE (4)
CLIMA (2)
TAMBURO (3)
PANTALONE (4)
CARTOLINA (4)
NEVE (2)
PIEDE (3)
RUOTA (3)
ORTO (2)
MARZIANO (4)
CARTONE (3)
COPERTA (3)
TRENO (2)
RAMO (2)
ASCENSORE (4)
MELONE (3)
ARTICOLO (4)
TAVOLO (3)
FINESTRA (3)
FORZA (2)
FIUME (3)
GAROFANO (4)
VETRO (2)
CAVALLO (3)
INCIDENTE (4)
MONTANARO (4)
DRAGO (2)

Grammatical Information Effects in Auditory Word Recognition*

L. Katz,** S. Boyce,† L. Goldstein,† and G. Lukatela††

Three lexical decision experiments were concerned with the separability of syntactic and semantic processing in spoken word perception. An additional experiment examined the problem of measuring reaction times to a spoken stimulus. Words in the Serbo-Croatian language were used; each stimulus consisted of a noun stem (which was either a meaningful root or a pseudoword stem) plus an inflectional suffix that conveyed information about the noun's grammatical case. Speed of identifying the inflectionally related forms of a noun was a function of differences in their syntactic meanings rather than differences in their physical forms or their actual frequencies of occurrence. In addition, identification of a noun was facilitated when it was preceded by a stimulus carrying predictive inflectional information whether that stimulus was a real adjective or pseudoadjective. The results echo previous findings for word perception in print and provide evidence of essential structural uniformity in the processing of inflection for both spoken and printed words. For both, there is evidence that inflectional processing is modular, at least to the extent that it is independent from semantic processing for the initial portion of its time course.

INTRODUCTION

The principal language of Yugoslavia, Serbo-Croatian, has provided a useful vehicle for psycholinguistic study. Its structure contrasts sharply with that of English, the language that has been used most often in such studies, and, so, provides a perspective that would otherwise be lacking, an approach that helps to separate universal principles from language-specific ones. The present studies, which use the Serbo-Croatian language, elaborate that perspective but are, additionally, concerned with a generality of a different kind: the extension of previous findings on the perception of printed words to the perception of spoken words.

Our intention was to test the cognitive generality of two previous findings that have implications for theories of lexical organization and the processing of grammatical information (Gurjanov, Lukatela, Moskovljević, Savić, & Turvey, 1985; Lukatela, Gligorijević, Kostić, & Turvey, 1980). We were concerned with the processing relation between the semantic and syntactic parts of a word as represented by the root of the word and its syntactic inflectional suffix. The primary question is whether these two sources of information are processed independently or not, at least for some significant portion of their time course. In brief, we wanted to assess the modularity of inflectional processing.

The group of Slavic languages, which includes Serbo-Croatian, depends on inflection as the major means of communicating grammatical information. In contrast, English depends strongly on word order to convey such information. For purposes of this discussion, it will suffice to describe, briefly, the Serbo-Croatian

system for the grammatical inflection of the noun word class. A grammatical inflection is added as a suffix to each noun stem to convey the information that the noun is, for example, the subject of the verb, or the object, the indirect object, an instrument, a location, etc. In addition, the inflection conveys information about number (is the noun singular or plural?). These different combinations of a fixed word stem plus an inflectional suffix are called the cases of the noun. For example, some cases for the word for 'woman' are *zena* ('woman' as subject of the verb: the nominative case), *zene* ('of the woman': genitive case), *zeni* ('to the woman': dative case), *zenu* ('woman' as object of the verb: accusative case), and *zenom* ('with the woman': instrumental case). The nominative case form is the base form in the sense that it is the citation form (e.g., the form used in dictionaries); the other case forms, as a group, are referred to as the oblique cases.

For the experimental psychologist, there is a certain tactical advantage to be found in using an inflected language to study the processing of syntactic information. The close proximity of (1) the lexical semantic information conveyed by the stem and (2) the grammatical information conveyed by the inflectional suffix, both packaged into a single word, makes it possible to study the processing of the the two kinds of information by means of standard word recognition paradigms such as the lexical decision task. We take advantage of the proximity of the two kinds of information in order to address the central question of this paper: Is the morphological distinction between inflectional and semantic information mirrored by independence in their mental processing? Or, in contrast, is inflectional analysis informed by semantics?

Inflectional morphology in Serbo-Croatian nouns and adjectives appears in the form of a word suffix; it gives the meaning of the word's case role in the phrase. It can be distinguished from other kinds of syntactic information such as noun gender and from other other kinds of morphology, e.g., derivational morphology (by which diminutives are formed, nominalizations are formed, etc.). The experiments in this paper are addressed specifically to the question of the mental processing of inflection. Because derivational morphology has a different function and is much less pervasive than inflection, it may well be subserved by a different processing system.

A seminal experiment on this topic has been reported by Lukatela et al. (1980). It was concerned with the process of printed word identification for inflectionally related nouns, viz., the different case forms of a given noun, e.g., *zena*, *zene*, *zeni*, etc. In that experiment, in spite of the fact that the various case forms have distinctly different frequencies of occurrence in the language (Dj. Kostić, 1965), reaction times for word identification in lexical decision experiments did not correlate with case frequency but, instead, fell into two groups. Reaction times were (1) identical among all the oblique case forms and (2) these were slower than the reaction time to the nominative case form. The equality of the oblique cases was surprising in light of the fact that one of these cases, the genitive, is used nearly as frequently as the nominative while another, the instrumental, occurs less than one-tenth as often. Thus, it was clear that the frequency of a noun's case form did not determine identification of that noun. Instead, it was the case form's syntactic identification (nominative or oblique) that affected the latency. Thus, Lukatela et al. found what appeared to be a processing difference that was a function of only syntactic structure; the mental machinery involved in word perception was apparently organized to process pure syntactic information. The term "satellite model" was applied to describe the pattern of reaction times because the nominative form was viewed as being central (the privileged form), surrounded by the slower, oblique, forms that were equal among themselves.

As suggested above, the only kind of explanation that can account for the superiority of the nominative case form is one that proposes an analysis of the word

form in which one component is specifically *syntactic* in nature. The nominative form can not consistently be differentiated from the other forms on the basis of superficial stimulus properties. For example, the nominative form zena can not be identified as a nominative form by reference to either the stem (*zen-*) or the inflection (*-a*); the former occurs for all of the other cases of zena and the latter occurs for all masculine and neuter genitive forms, all neuter accusatives and some masculine accusatives, in addition to the feminine nominative. Therefore, in order to identify the case of a noun, its inflection must be interpreted in the context of the "gender" (i.e., declension) of the noun, information that is more or less arbitrary and is available only as part of its lexical entry. Thus, no analysis of a stimulus into its phonological or graphemic properties provides, by itself, an identification of the stimulus as a nominative form and, therefore, no such analysis can provide the nominative form with privileged lexical access.

The critical assumption in this analysis is that stem and inflection are encoded by separate processes: the stimulus must be analyzed into distinct semantic and syntactic components. The syntactic information required for noun identification includes word class (e.g., that the word is a noun and not a verb, adjective, etc.), the noun gender and, finally, the noun's case role in the phrase. Analysis of the inflection must first be preceded by information about word class and gender. Although the present research does not pursue the question of the locus of the gender information, it seems plausible to consider it to be stored in the lexicon along with the semantic (nonsyntactic) meaning of the stem. Processing of the meaning of the suffix (i.e., the noun's case role) may then proceed without interacting with the semantic meaning of the stem. Therefore, this explanation supposes that there are, in fact, separate semantic and inflectional processors.

There is additional experimental evidence to support this notion of separable semantic (lexical) and syntactic (inflectional) processors. Gurjanov et al. (1985) visually presented, on each trial, an adjective followed by a noun; a lexical decision was required on only the noun. Gurjanov et al. were looking for an effect of the syntactic relation between adjective and noun on the perception of the noun. In Serbo-Croatian, the inflections on a noun and the adjective modifying it must agree in gender, case, and number. However, such agreement does not entail phonological identity; these suffixes on the adjective and noun do not necessarily sound (or look) the same. For example, the adjective-noun pair *vitkoj guski* ('to the slim goose') has the congruent inflections *-oj* and *-i*. Thus, agreement (i.e., congruency) exists only at the level of syntax; not at the level of phonology. (This is a not uncommon situation in the languages of the world.) Table 4 presents some examples of Serbo-Croatian adjectives and nouns.

Not surprisingly, when subjects in the Gurjanov et al. (1985) study were presented with adjective-noun pairs in which grammatical agreement occurred, subjects were faster in making a lexical decision than when the members of the pair were incongruent (differing in case). More interestingly, the relative facilitation produced by grammatical congruency occurred even when the stimulus preceding the noun was a pseudoadjective, that is, a meaningless nonword stem with a real, appropriate adjectival suffix. This apparent effect on the perception of a noun that followed a pseudo- adjective could only have occurred by linking the inflections (the syntactic information code) of the pseudoadjective and noun. The congruency effect of pseudoadjective primes is a second piece of evidence (in addition to the satellite pattern reported by Lukatela et al., 1980) that suggests that words are analyzed into distinct semantic and syntactic codes and that the process of assigning meaning to an inflection is independent of the semantic status of the stem. Such grammatical congruency effects have been observed in Serbo-Croatian in other situations, e.g., pronoun-verb combinations (Lukatela, Moraca, Stojnov, Savić, Katz, & Turvey,

1982), and preposition-noun combinations (Lukatela, Kostić, Feldman, & Turvey, 1983). Thus, there is consistent evidence of the kind that is needed to support the suggestion of the Lukatela et al. (1980) studies, that there is psychological reality to the notion of modular inflectional processing, at least for perception of the printed word.

A specific purpose of the present research was to determine if a syntactic congruency effect could also be found with auditory presentation. In fact, there was some reason to believe that the effect found with print might be due merely to an artifact of that particular presentation mode and would not be found when speech stimuli were used. As is common in the visual presentation paradigm, the entire stimulus, word or nonword, consisting of stem plus inflectional suffix had been displayed simultaneously in the Gurjanov et al. (1985) study. With simultaneous presentation, subjects could have easily attended to the inflection before the stem or in parallel with it. It could have been this peculiar strategy of focusing attention that was responsible for the apparently independent use of inflectional information. With speech, on the other hand, such an attentional strategy is much less feasible. Because the stimulus develops in time, the stem is completed before the onset of the inflection. Therefore, the subject must perceive the stem first, at least at some level of perception.

The present studies replicated two of the key experiments discussed above, using speech stimuli instead of print, in an attempt to assess the validity and generality of the previous results and, if possible, to advance their interpretation. The main experiments we present are (1) a simple lexical decision experiment designed to look for RT differences between nominative and oblique case nouns (the auditory analogue of the satellite pattern found by Lukatela et al., 1980) and (2) two adjective-noun priming experiments designed to look for auditory pseudoadjective syntactic congruency effects (an analogue and an extension of the Gurjanov et al., 1985, study).

However, experiments on speech introduce methodological problems not present in print experiments, due to the temporally dynamic nature of speech. In the typical print lexical decision paradigm, in which all parts of the stimulus are presented simultaneously, it is reasonable to begin the reaction time clock at the point when the stimulus appears on the screen. Spoken words, on the other hand, are presented continuously over a period of time (250-1000 ms) that is long relative to the latencies of responses to those stimuli. During this time, the listener may be making successive judgments about what he or she has heard so far. Thus, it is reasonable to ask at what point during the temporal unfolding of the word the timing of subjects' responses should start.

There is evidence from mispronunciation detection and short-latency shadowing studies that listeners make use of information in the unfolding word during the course of its presentation (Cole & Jakimik, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978); often they are able to identify the word before hearing the full extent of the stimulus. Marslen-Wilson and his colleagues have argued further that word recognition in real time is a matter of selecting the correct word from among a field of possible choices; when the first few segments (roughly the first two phonemes) of the word have been scanned, a list of candidate words (called a "cohort") with the same initial phonemes is called up from the lexicon. As the succeeding segments are identified, they serve to narrow down the list of candidates until only one word remains as a possible match to the input. Word recognition is achieved at this point. Thus, in the absence of syntactic/semantic context (which Marslen-Wilson and his associates claim further narrows down the field of candidave words), recognition of any given word occurs at a stable point in the left-to-right sequence of phones, which point is determined by the intersection of the individual phonological properties of the stimulus word and the properties of other words in the lexicon.

However, even though lexical access may be achieved before the word ends, the lexical decision response itself may be delayed further. First, there may be uncertainty that the word has, in fact, ended; some accessed lexical items may be embedded within another item (for example, *candid* in *candidate* and *iskren*, 'sincere' in *iskrenost*, 'sincerity'). Second, the response may be obligatorily delayed until after the inflectional suffix.

In the present experiments, we were not concerned with the factors that affect the identification points of different noun stems. Rather, we were interested in distinctions among the various case forms of a single noun stem, i.e., syntactic inflectional effects. The technical problem facing us was to be able to unambiguously interpret reaction time differences among the various case forms as being due to syntactic factors and not to phonological factors. For example, lexical decision time to the nominative form *zena* might be faster than reaction time to the instrumental form *zenom* either for a theoretically interesting reason (e.g., because nominatives have privileged lexical access) or for an artifactual reason (e.g., the instrumental form contains an additional phoneme that might lengthen the lexical decision process). Even when the oblique case forms do not contain more phonemes than the nominative, a similar argument could be made on the hypothesis that their different phonetic shapes may take longer to process.

Experiments 1 and 2 attempt to deal directly with this issue of the measurement point in the auditory modality by comparing the results obtained from presentations using different measurement points. These results are, in turn, compared to the print experiment data in Lukatela et al. (1980).

EXPERIMENT 1

Experiment 1 was designed to look for faster identification of spoken words in their nominative case form than in their oblique forms. As in the Lukatela et al. (1980) study, the nominative case was compared with two oblique cases, the genitive and instrumental. Additionally, the experiment went beyond the Lukatela et al. print study by including an explicit contrast between high and low frequency words. While the frequency effect is well established for lexical decision in the visual mode, it is not predicted per se by the Marslen-Wilson model (e.g., Marslen-Wilson & Welsh, 1980), and it seemed appropriate to test it in auditory mode. Experiment 1 was also designed to test the prediction that nonword stimuli that resemble words for a greater (left-to-right) temporal portion of their length will show longer RTs than stimuli that become nonwords early on.

Method

Stimuli

Words. Twenty-four high frequency and 24 low frequency nouns were selected from the word frequency dictionary prepared by Dj. Kostić (1965) from a corpus of approximately 14,000 words sampled from three years of the daily newspaper *Politika* (approx. 1953-1955). Of each high and low frequency group, half were masculine and half were feminine gender nouns. High frequency words were defined as occurring more than 32 times in the Kostić corpus. Low frequency words occurred once in the the same corpus.

Two native speakers independently checked the words selected for (1) current usage; (2) stress on the initial syllable and consistent accent pattern throughout the three cases; and (3) invariant meaning over grammatical cases. Each noun contained a regularly inflected 2-syllable stem.

The cases used were nominative, genitive, and instrumental singular. In all forms except masculine nominative there was an additional syllable corresponding to the case ending. Words were not always highly concrete in meaning, although words with obscure or literary denotations were avoided.

Pseudowords. Forty-eight pseudowords were constructed to be phonotactically possible but non-occurring words in Serbo-Croatian, and to become pseudowords at one of four possible points from left to right in a 2-syllable stem. This was done by finding non-occurring initial sequences of varying lengths in a Serbo-Croatian dictionary (Grujić, 1969) and adding enough material to fill out a 2-syllable stem. For example, the pseudoword *bemaz* was made by first discovering that words beginning *bel-* and *ben-* existed, but none beginning *bem-*, and then adding the final sequence *-az*. The four deviation points were defined as (1) at the first vowel; (2) at the middle consonant or consonant cluster; (3) at the second vowel; (4) at the final consonant or consonant cluster of the stem. The words were checked by two native speakers for acceptability and resemblance to real words--in particular, we tried to avoid pseudowords that differed from an actually occurring word by only one phonetic segment.

The practice list contained pseudowords with each of the four possible deviation points. Thus, subjects were trained to avoid making an incorrect 'word' decision based on the early word-like portion of the pseudoword.

Stimulus tape. The list of words was read three times by a male native speaker of the Belgrade dialect in a sound-treated booth. To avoid the list-like intonation that might have occurred if all three cases had been read together, the words were read in blocks of all the nominatives, followed by all the genitives, and all the instrumental case words. The best and most intelligible recording for each word, as chosen by a second native speaker, was then digitized at a sampling rate of 20 kHz using the Haskins Laboratories' Pulse Code Modulation system. The beginning of each word was located automatically by the amplitude of its onset. In a few cases where the amplitude was too low, the onset was adjusted manually. The digitized speech file of stimuli was then randomized and recorded onto tape. A 20 ms marker pulse, timed to coincide with the beginning of the word, was recorded onto a second channel; this was used to start the RT clock.

The tape consisted of a practice list of 35 words and pseudowords followed by the main test list of 102 words divided into 6 lead-in (dummy) words and 96 test words. The stimuli were recorded with an interstimulus interval of 5 seconds.

Design

Each subject heard a total of 96 words in the main test, of which 48 were words and 48 were pseudowords. No subject heard any given noun more than once in the course of the test. This was achieved by dividing the 96 words into three groups (A,B,C) and the 45 subjects into three groups (I, II, III). Each group of 16 real words contained 8 high frequency and 8 low frequency nouns, of which half again were masculine and half feminine gender. Subjects in Group I saw words from Group A in the nominative case, from Group B in the genitive case, and from Group C in the instrumental case. For subjects in Group II the categories were B,C,A respectively, and for subjects in Group III the categories were C,A,B. Nonwords were partitioned similarly into categories and subject groups, with the exception that instead of a two-level division into high and low frequency items, they were divided into four categories based on constructed deviation point. Thus, each group of 16 consisted of 4 groups of 4, each with a different deviation point. Of each group, half again were controls for the masculine gender suffixes and half were controls for the feminine gender. Note that the suffix inflections on pseudowords are quite ambiguous with regard to specification of case because interpretation of a suffix must, in part, depend on the

lexically specified knowledge about the noun's gender. For example, the suffix -a on a pseudoword could be interpreted as either the nominative singular case (if the stem is taken to be a feminine gender word) or the genitive singular (if the stem is taken to be a masculine gender word). The reason for manipulating suffixes on pseudowords was in order to balance the number of presentations of each real suffix.

Procedure

The stimulus tape was played to the subjects from an Uher tape recorder over headphones. The headphones had a cutoff frequency of 8000 Hz. On each trial, the subject's task was to decide as rapidly as possible whether the auditory stimulus was a word or not. Both hands were employed in responding to the stimulus. Both thumbs were placed on a telegraph key close to the subject, and both forefingers were placed on another telegraph key 4.5 cm farther away. The closer button was depressed for a "no" response (the auditory stimulus did not correspond to a word) and the farther button was depressed from a "yes" response (the auditory stimulus corresponded to a word).

During the interstimulus interval, the subject was required to write down 'sigurna' (sure) or 'non sigurna' (not sure) as an indication of confidence about the accuracy of each response. There was ample time for the subject to be ready for the next response. The entire session lasted approximately 30 minutes with a short break after the practice session and after 51 items in the main test.

Subjects

Subjects were 45 undergraduate students at the University of Belgrade who participated as part of their course requirements. All had had recent experience with reaction time experiments in the visual mode but little or no experience with auditory experiments. All were speakers of the Belgrade dialect.

Results and Discussion

Words

Errors of classification (word-pseudoword) and all latencies less than 300 and greater than 2000 ms were excluded from the data set. Error rates were 1.5% for words, 2.5% for pseudowords, and were unsystematic. Separate analyses of variance were performed using, respectively, subject and stimulus item variability as the error term. Data from the three subject groups were pooled after a simple one-way analysis of variance failed to show any significant difference between them. For the subjects analysis, means were calculated over the four items that appeared in each combination of Frequency, Gender, and Case; for the items analysis, means were calculated over the 15 subjects who saw that item in a particular case. A priori, we had expected a main effect of frequency and a main effect of case. It was further predicted that the case effect would break down into a pattern in which the oblique cases would show longer RTs than the nominative but would not differ from each other.

Table 1 presents mean RT for each combination of Frequency (High/Low), Gender (Masculine/Feminine) and Case (Nominative/Genitive/Instrumental). Both of the predicted main effects were significant. For Frequency: subjects, $F(1,44) = 204.26$, $MSe = 4873$, $p < .001$, and items, $F(1,44) = 22.61$, $MSe = 13632$, $p < .001$. For Case: subjects, $F(2,88) = 16.58$, $MSe = 6605$, $p < .001$, and items, $F(2,88) = 5.31$, $MSe = 4725$, $p < .01$. There were no other significant effects. The essential results of Table 1 are presented in Figure1.

Orthogonal contrasts indicated that the two oblique cases did not differ between themselves and that a large proportion of the significant case effect was due to the difference between nominative and oblique cases. The contrast Nominative vs.

Genitive-Instrumental gave, for the subjects analysis, $F(1,44) = 32.9$, $p < .001$, and for the items analysis, $F(1,44) = 14.48$, $p < .001$. The contrast Genitive vs. Instrumental was, for subjects, $F(1,44) = .55$, n.s., and for items, $F(1,44) = .12$, n.s.

TABLE 1

Mean Reaction Times and Standard Deviations for Real Words in Experiment 1 as a Function of Case, Frequency, and Gender.

		CASE					
Freq	*Gender*	*Nominative*		*Genitive*		*Instrumental*	
High	Masculine	742	47	779	53	780	60
	Feminine	714	66	786	101	757	69
Low	Masculine	854	105	882	144	843	116
	Feminine	809	75	840	70	885	92
Mean		780	92	822	104	816	98

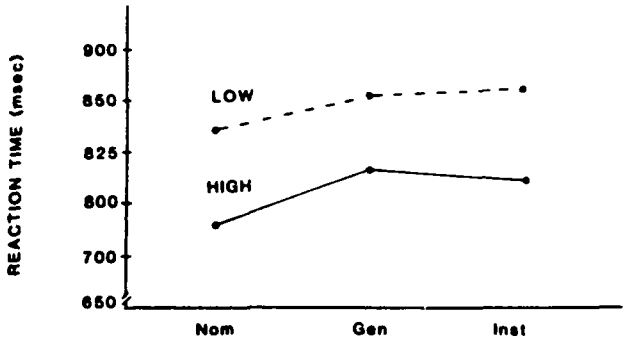


Figure 1. Lexical decision reaction time to real nouns preceded by real or pseudoadjectives with congruent or incongruent case matches.

These results correspond reassuringly with those of Lukatela et al. (1980). Overall, however, latencies to our auditory stimuli were longer than latencies to the visually presented stimuli of Lukatela et al.

This experiment was also designed to test the 'left-to-right' processing model (or cohort model) of auditory word recognition by using pseudowords that resembled existing words up to various divergence points in their temporal left-to-right sequence. The prediction of the cohort model was that pseudowords with earlier points of divergence would be rejected faster.

As was the case for real words, separate analyses of variance using subject and item variability were carried out for nonwords. Factors here were Position of Divergence (Points 1 to 4), Gender Control (Masculine/Feminine), and Case Control (Nominative/Genitive/Instrumental). (Recall that the Case and Gender factors for pseudowords are only methodological controls designed to balance the number of occurrences of the suffixes that occurred with real words.) Table 2 presents the mean reaction time for each combination of factors. In the subjects analysis, all main effects and interactions except the three-way interaction were significant. On the

other hand, in the items analysis only Case was significant, and that marginally, $F(2,80) = 3.34$, MSe = 4448, $p < .05$.

TABLE 2

Mean Reaction Times and Standard Deviations for Pseudowords In Experiment 1 as a Function of Case, Gender, and Position of the Divergence Point.

		CASE					
Position	*Gender*	*Nominative*		*Genitive*		*Instrumental*	
One	Masculine	851	110	943	164	938	180
	Feminine	776	114	804	154	812	148
Two	Masculine	831	172	866	142	868	136
	Feminine	885	162	884	166	854	162
Three	Masculine	876	136	907	186	921	189
	Feminine	834	161	879	159	860	153
Four	Masculine	915	130	889	156	963	161
	Feminine	900	176	899	170	926	166
Mean		858	151	884	165	893	168

This weak outcome for the effect of position of the divergence point offers no useful evaluation of the notion that the process of word recognition involves a reductive winnowing of candidate words within a cohort. However, the relatively greater strength of the case effect suggests that if a reductive process exists, the time scale of the process may be substantially shorter than the time consumed by inflectional processing.

RT Measurement Considerations

One might conclude from these results that lexical decision making is the same for both the auditory and visual modes. However, there are aspects of response timing for the respective modes that must be considered before such an interpretation is warranted for real words. The results of these analyses were based on the raw RT as measured from the beginning of the word. As mentioned above, a serious consideration, when comparing different spoken words, is whether to make the starting point for a lexical decision latency an arbitrary and uniform point for every word (e.g., starting each measure from the beginning of the word) or to make it a different point for every word, based on the properties of the individual word. However, for Experiment 1, the important comparisons are not the typical ones made between different nouns (i.e., between different stems); rather, the relevant comparisons are between different case forms of the same noun stem. Therefore, in order to interpret appropriately the reaction times to these different inflected forms, we must know the point in a speech stimulus where that form can be discriminated from the other cases. The following experiment was designed to discover the perceptual identification point for each inflection in each noun in the present experiment. Having found this point, we can then reanalyze the raw reaction time data but with RT scores adjusted individually for each stimulus item. For example, we wished to make accurate comparisons of lexical decision latencies to the nominative *zena*, the genitive *zene*, and the instrumental *zenom*, based on the individual identification points of each.

EXPERIMENT 2

This experiment used a variation of Grosjean's (1980) 'gating' procedure. In the original experiment, subjects were presented auditorily with successively longer sections of the stimulus word and asked to guess what the word might be. They were also asked to indicate on a rating scale how confident they were about their choice. This procedure made it possible for Grosjean to monitor which words were being entertained as candidates by the subjects at any point during the word, as well as the point at which they began to identify the word consistently.

Method

Our concerns were slightly different from Grosjean's (1980). We wished to determine the point in a given suffix-inflected word at which the item diverges from other inflected forms with the same stem; that is, the point where the different case forms of the same lexical item are perceptually discriminated. Armed with these discrimination points, we could then calculate (for the data in Experiment 1) an adjusted lexical decision RT to each of the different case forms for each stem, clocking each RT from the discrimination point for its stimulus.

Subjects were presented with successively longer left-to-right sections of the stimulus item, but for each stimulus trial the subject was presented with a list of the responses he or she was allowed to choose from. The gating interval was set at 20 ms. Thus, on each successive trial the section of the stimulus word presented was 20 ms longer, measured from the beginning of the word, than in the previous trial. The subject was given four alternatives to choose from for each trial. Three of the possible responses were the nominative, genitive, and instrumental case forms of the lexical item in question. Data from these were designed to identify the point at which a response to the case form might be initiated. Additionally, there was a fourth possible response that corresponded to the 2-syllable stem form of the word plus an extra syllable beginning with a stop consonant. This condition was inserted to provide a pseudoword control for the nominative case of masculine nouns, which has zero marking. On a phoneme-by-phoneme basis, the three cases for nouns in Serbo-Croatian diverge at similar points; thus *sadnica* (seedling), a feminine noun in the nominative, differs from its genitive form *sadnice* and its instrumental form *sadnicom* in the vowel following /c/. (On an acoustic/perceptual basis, this is less clear, of course; differing amounts of the vowels /a/, /e/, and /o/ may be required for identification in any given phonetic environment, and anticipatory nasalization from the /m/ may be present.) The situation is similar for masculine nouns; *prozor* (window) in the nominative differs from *prozora* and *prozorom* in what follows /r/; however, in the masculine nominative, what follows is silence. We felt we needed some way of determining for these stimuli at what point subjects would feel confident that there was no phonological material following the stem. Stop consonants were chosen to begin the extra syllable so that we could test subjects' ability to distinguish the condition of a null inflection on a masculine nominative noun from the onset of a syllable that begins with silence.

Four consonants, /t/,/d/,/k/, and /g/, were used in conjunction with the vowel /a/. These were concatenated to the 2-syllable word stem in accordance with a Serbo-Croatian constraint against dissimilar voicing in clusters. Geminates were avoided. Thus, for example, the stem *fabrik* received the pseudoword control *fabrikta*, *nastup* received *nastupta*, and *prinud* received *prinudga*.

The 48 masculine and feminine real word nouns used in Experiment 1 were divided into four sets of 12, of which half were feminine and half were masculine nouns. There were 32 subjects divided into four groups of 8. Each group heard gating trials for one set of nouns in the nominative case, one set of different nouns in the genitive

case, one set of different nouns in the instrumental case, and a fourth set of different nouns with an extra syllable at the end. Some of the subjects had previously participated in Experiment 1 but to minimize familiarity effects several days elapsed between experiments.

Because of the large number of gating intervals involved when starting from the beginning of the word (some longer words had as many as twenty 20 ms gating trials) and the fact that the choice point between possible responses came late rather than early in the stimulus, we did not begin presentations with only the first 20 ms segment. Instead, we first submitted the gated stimuli for judgment to a panel of four native speakers, with instructions to indicate the first trial at which they detected any hint of the identity of the correct inflection. (The judges were told which word stem to expect on each trial.) Gating trials for the 32 subjects in the experiment proper were then set to begin two gates (40 ms) before the earliest point indicated by any judge. The gating tapes were played over headphones, with 1 second between trials and approximately 5 seconds between trials for different stimuli.

Results and Discussion

Subject answer sheets contained, for each trial, the list of four possible responses, on which subjects were instructed to mark their choice for identification of the stimulus, and a continuous scale from 0 to 100, on which subjects were instructed to indicate how confident they felt about that choice. The first trial on which a subject indicated he or she was 80% or more confident of identification for the stimulus was recorded as the recognition trial for that subject. However, if the subject later changed his or her identification or indicated a drop in confidence level, the next 80% point was chosen. The eight recognition trials so obtained for each stimulus were averaged, and converted to units in milliseconds from the beginning of the word. For example, if the various subjects' choices for 80% identification of the stimulus *nacinom* (masculine, instrumental case) were trials 7,7,8,6,7,8, and 10, and the first trial began 200 ms into the stimulus, the recognition (i.e., identification) point was defined as 200 plus 20 times the mean of the eight trial choices, or 200 + (20 x 7.63) = 352.6 ms from the beginning of the word.

These steps produced a matrix of recognition point values corresponding to the set of different nouns, with each noun in three cases, used in Experiment 1. These values were then used in an analysis of covariance as covariate control scores (the dependent variable being the same as before, i.e., raw lexical decision latency). Thus, the analysis of covariance was used to test the satellite pattern on the adjusted means. Because subjects for the two experiments differed, only an items analysis of covariance was possible. Strikingly, the pattern of results was nearly identical to that obtained for the analysis of variance using raw reaction time scores only. Frequency was significant, $F(1,43) = 27.48$, MSe = 9839, $p < .001$, as was Case, $F(2,87) = 5.01$, MSe = 4778, $p < .01$; there were no significant interactions. The covariate itself was significant between-items, $F(1,43) = 17.97$, MSe = 9839, $p < .001$, correlation coefficient = .51, but not within Case. Orthogonal contrasts on the Case effect again revealed a pattern in which the nominative was faster than the other two cases, but the oblique cases did not differ between themselves (Nominative vs. Genitive-Instrumental: $F(1,43) = 14.06$, $p < .001$; Genitive vs. Instrumental: $F(1,43) = .54$, n.s.).

We also performed an analysis of variance on just the recognition point data for the three cases. Cell means for these data are reported in Table 3. The analysis revealed no effect for Frequency or Gender but a significant effect did occur for Case, $F(2,28) = 9.82$, MSe = 1631, $p < .001$, and for the Case x Gender interaction, $F(2,88) = 3.21$, MSe = 1631, $p < .05$. This outcome appeared to stem from systematic differences in length among the cases; instrumentals are longer for both genders and nominatives are shorter than genitives for the masculine nouns. Thus, the perceptual

recognition point was apparently sensitive to length of the suffix itself. However, when this length difference among different cases was statistically corrected (in the analysis of covariance reported above), reaction times still followed a satellite pattern.

TABLE 3

Mean Gated Recognition Points and Standard Deviations for Real Words in Experiment 2.

		CASE					
Freq	*Gender*	*Nominative*		*Genitive*		*Instrumental*	
High	Masculine	515	92	525	63	554	60
	Feminine	503	76	471	69	522	55
Low	Masculine	494	107	514	64	507	68
	Feminine	527	88	530	77	583	106
Mean		510	89	510	70	542	78

We also examined the recognition point results for those stimuli with an extra syllable that turned them into pseudowords (e.g., the pseudoword, *fabrikta*, from the real word, *fabrik*). Although on a phoneme-by-phoneme count these stimuli were no longer than nouns in the instrumental case, recognition points to these stimuli were typically from 20 to 60 ms later, averaging 50 ms later than the instrumental case, and later than the recognition points for the nominative case by 80 ms. It is clear, therefore, that the recognition of normal masculine nominative case words (which lack any inflectional suffix) is not delayed by a presumed uncertainty about when the word ended; those normal nouns, in fact, had earlier recognition points than longer items. Note, however, that because these extra-syllable words become phonemically identifiable as nonwords at exactly the same point as the nominative, genitive, and instrumental case real words, their longer response times must indicate some effect of their anomalous syntactic status. Nor can length alone be a sufficient explanation, because the extra-syllable is equivalent in length to the real words' instrumental case suffix.

All our analyses point to the same pattern; namely, that there is a stable reaction time advantage for the nominative case. We conclude therefore that the so-called satellite pattern is a robust aspect of the lexical decision process. Further, it seems that the choice of a measurement point is not crucial to tests of auditory lexical decision time for inflected words, as long as that measure is consistent over the different items.

EXPERIMENT 3

In Experiment 3, we studied the processing of noun inflections in adjective-noun pairs. Both the purpose and design of Experiment 3 were similar to those of the print experiment by Gurjanov et al. (1985) with the single exception that the modality of presentation in the present experiment was auditory. The Gurjanov et al. results had demonstrated that there was independent processing of inflectional information in the course of recognizing a word. As we discussed in the introduction to this paper, there was some reason to suspect that this apparent influence had been merely an artifact of the visual presentation mode. In visual presentation, it is possible for subjects to attend to the inflectional suffix before attending to the stem of the noun. In contrast this strategy is much less likely to occur when the stimulus is spoken

because the stem precedes the suffix temporally. Thus, the purpose of Experiment 3 was to determine whether or not grammatical congruency between an adjective and a noun--and particularly between a pseudoadjective and a noun--would affect the time to decide on the lexicality of the noun when the stimuli were auditorily presented.

Method

Design

An auditory lexical decision task was used in a priming paradigm. On each trial, the noun target was preceded by an adjective. However, the adjective was never associatively related to the noun. Noun and adjective combinations were selected on the basis of results of an earlier study (see Gurjanov et al., 1985) that asked subjects for associations to the adjectives used in the present study. The adjective-noun combinations used in the present study were those that had never been produced by subjects. Thus, the so-called adjective prime was unlikely to be an effective semantic prime.

Although the adjective was not semantically predictive, it did contain information enabling prediction of the inflection on the subsequent noun for half of the trials. The inflection on the adjective member of each pair either agreed, in case and number, with the inflection on the following noun or it disagreed in case (but not in number). Two cases were used: the nominative (agent case) and the dative (indirect object case). For the dative case, the masculine-neuter adjectival inflection was unique (for adjective inflections although not unique if noun inflections are included) and the feminine adjectival inflection was universally unique, so that the case of the dative case adjectives could be identified easily. In contrast, the nominative case suffixes were ambiguously either nominative, accusative, or vocative case for singular adjectives and could also be read as certain plural forms (except for the singular neuter form). The variable of case agreement between adjective and noun (grammatical congruency) was the major variable of the experiment. In addition, half of the adjectives were real adjectives known to the subject (that is, they had meaningful stems with real inflections) while half of the adjectives were pseudoadjectives, composed of pseudostems, but with real inflections. Finally, in accordance with the lexical decision paradigm, half of the noun targets were real words, while half were pseudowords, composed of a pseudostem, but with a real inflection. Pseudoword adjectives and nouns were phonologically legal. They were constructed by changing a consonant and vowel at random positions in a real adjective or noun. Pseudowords were the same items used in the Gurjanov et al. (1985) print experiment. Thus, all words and pseudowords in the experiment had real inflections. The variables were combined factorially to produce a 2 x 2 x 2 x 2 design, i.e., Adjective Lexicality (Word/Pseudoword) x Noun Lexicality (Word/Pseudo-word) x Noun Case (Nominative/Dative) x Grammatical Congruency (Agreement/No Agreement).

Each subject heard seven different examples in each of the 16 combinations of the four factors, for a total of 112 trials. No subject heard the stem of any adjective, noun, or pseudoword more than once. There were four groups of 16 subjects each. Among the four groups, each adjective-noun combination was given inflected suffixes that produced adjective-noun agreement once when the noun was in the nominative case and once when it was in the dative case. Likewise, there was adjective-noun incongruency once when the noun was in the nominative case and once when it was in the dative case (see Table 4). A practice list of 28 trials preceded the experimental list.

TABLE 4

Design Summary: Grammatical Congruence for Adjective-Noun

Noun Stem	Adjec Stem	Noun Case Inflection	Adjec Case Congruence	Example Adjec-Noun	
Real	Real	Nominative	Agree w/Noun	mladi	slon
			Not Agree	mladom	slon
		Dative	Agree	mladom	slonu
			Not Agree	mladi	slonu
	Pseudo	Nominative	Agree	bleti	covek
			Not Agree	bletom	covek
		Dative	Agree	bletom	coveku
			Not Agree	bleti	coveku
Pseudo	Real	Nominative	Agree	dragi	nod
			Not agree	dragom	nod
		Dative	Agree	dragom	nodu
			Not Agree	dragi	nodu
	Pseudo	Nominative	Agree	mafi	pavnot
			Not agree	mafom	pavnot
		Dative	Agree	mafom	pavnotu
			Not Agree	mafi	pavnotu

mladi slon: young elephant; bleti covek: -- man; dragi nod: dear --; mafi pavnot:----

Stimuli

A female speaker of the Belgrade dialect produced a token for each word and pseudoword in each case. When the words were produced, the adjectives were spoken one after the other, as in a list. Likewise, the nouns were spoken in a listwise manner. The tokens were digitized at 20 kHz and prime-target combinations were constructed from the digitized tokens. For example, the exact same digitized noun token (e.g., *slonu*) was used following the nominative case adjective, *mladi*, and the dative case form, *mladom*. That is, the same utterance was the target stimulus for both adjective-noun agreement and disagreement. Thus, no information relating to the prosodic contour of a normal adjective-noun utterance existed in the synthesized pairing. The silent interval between each adjective and noun was approximately 630 ms. Following the results of Experiment 2, a subject's reaction time to a noun target was measured from a marker pulse signaling the onset of the target. The interval between trials was 2 sec.

Subjects

Subjects were 64 psychology students at the University of Belgrade who were required to participate as part of a course requirement.

Results and Discussion

The question under study was: Would spoken grammatical information, in the form of inflectional congruence between an adjective and a subsequent noun, affect identification of the noun? Further, if recognition of a congruent target noun is facilitated relative to an incongruent noun, is this facilitation effect just as large when the preceding stimulus is a pseudoadjective as when it is a real adjective? That is, will appropriate inflectional information be as effective in a stimulus that has no lexical/semantic meaning as in a stimulus that does have such meaning? Or, on the other hand, is the processing of inflectional information inextricably bound to the

processing of the stem; is there no processing of the inflection that is distinct from semantic processing?

Figure 2 presents the average reaction times for recognition of real noun targets. Error rates were low (1%-3%) and were unsystematic. Reaction time is plotted as a function of the grammatical congruence of the adjective-noun combination (i.e., agreement or no agreement in their case inflections). It also shows the lexical status of the stem of the priming stimulus (i.e., whether or not it was a real or pseudo-adjective stem). The data have been averaged over the nominative and dative cases. The results show a large effect of congruence; when both members of the pair agree in case, reaction time is faster. Moreover, this facilitation appears even when the priming stimulus is a pseudo- adjective and not only when it is a real adjective. Nevertheless, the facilitation effect appears to be stronger when the priming stimulus is a real adjective.

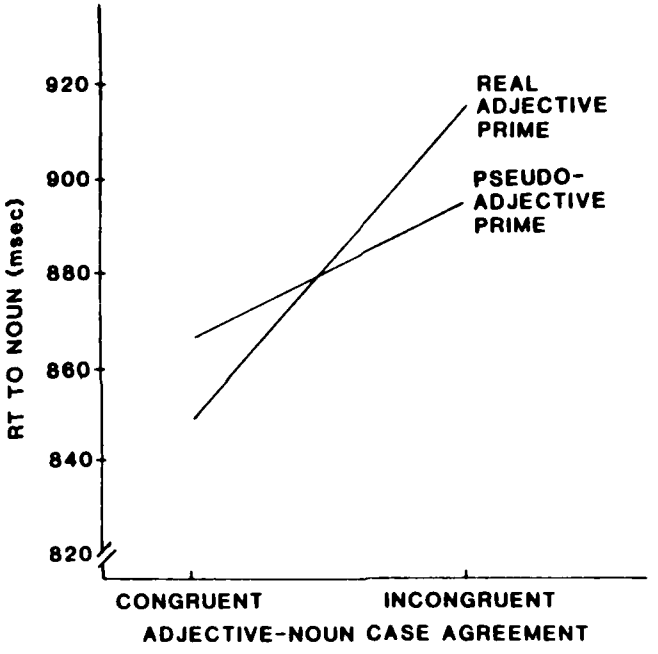


Figure 2. Lexical decision reaction time to real nouns prceded by real or pseudoadjectives with congruent or incongruent case matches and long (800 ms) or short (zero) interstimulus intervals between adjective and noun.

All of these effects were confirmed by analyses of variance. For the main effect of Grammatical Congruency, the subjects analysis of variance produced $F(1,60) = 67.44$, MSe = 4691, $p < .001$, and the items analysis of variance produced $F(1,54) = 42.75$, MSe = 3254, $p < .001$. In the items analysis, Grammatical Congruency accounted for 50.3% of the total variance. For the interaction between Grammatical Congruency x Adjective Lexicality, the subjects analysis produced $F(1,60) = 12.7$, MSe = 4691, $p < .001$, and the items analysis produced $F(1,54) = 6.89$, MSe = 3254, $p < .02$. In the latter analysis, the interaction accounted for 8% of the total variance. Thus, although there was an interaction, the main effect of Grammatical Congruency was stronger, accounting for over four times as much of the variance in RT. The only remaining significant result, not shown in Figure 1, is that of Noun Case. Nouns in the

nominative case were recognized faster than nouns in the dative case (a result that appears consistently in all of our experiments). This main effect was significant both in the subjects analysis of variance, $F(1,60) = 57.61$, MSe = 4033, $p < .001$, and in the items analysis of variance, $F(1,54) = 10.08$, MSe = 10168, $p < .003$.

The pseudonoun data were inspected and no consistencies appeared except that pseudonoun reaction times ("Nonword" responses) were slower than responses to real nouns. The lack of consistency was not surprising because most of the pseudonoun suffixes we used were uninterpretable given the absence of a lexical entry for the pseudonoun and, therefore, the absence of information about the target's gender (i.e., declension). The suffixes that were used were mainly those that take on different case meanings depending on the gender of the stem. For example, the suffix -u would be interpreted as having dative case meaning for a real masculine or neuter noun but would be interpreted as accusative case for a real feminine gender noun. Note that there is no comparable problem with pseudoadjective stimuli. Suffixes are often ambiguous for real adjectives (disambiguation must await the appearance of its noun complement). Moreover, for the real and pseudoadjective suffixes used in the present experiment, only one, -i, was ambiguous.

The overall results of Experiment 3 are similar to those found in the print experiment of Gurjanov et al. (1985), with one exception In print, no congruency effect was found for nominative case nouns; only the dative case nouns (and, in a second experiment, genitive case forms) were affected by grammatical agreement or disagreement with a preceding adjective.

Experiment 3 was informative in two ways. First, note that processing the inflection appears to be an obligatory process; congruency effects were found even though the stem was heard before the inflection and even though it was not necessary for the subject to attend at all to the inflection--which was always a real inflection and was, therefore, irrelevant to the subject's decision about the lexical status of the stimulus item (i.e., whether it was a word or pseudoword). Second, the answer to our question about the independence of syntactic processing from lexical/semantic processing is less than straightforward, but it is informative, nevertheless. It is clear that in auditory presentation as well as in print there is a substantial facilitating effect of congruence between grammatical markers. This facilitation is strong even when the priming stimulus is not a real word. The latter point is particularly important because it suggests that there is a means by which syntactic processing can proceed without semantic support. However, it is also the case that the facilitating effect of congruence is even greater when the priming stimulus is, in fact, a real word and has a lexical representation. This interaction leaves us with an ambiguous outcome with regard to the hypothesis that inflectional information is processed without any reference at all to semantic information.

EXPERIMENT 4

The fourth experiment was designed to address the question of the naturalness of the congruency facilitation effect found in Experiment 3. In Experiment 3, the interstimulus interval (the time between the end of the adjective and the onset of the noun) was approximately 630 ms. This is much larger than the interval in natural connected speech. In fact, in rapid normal speech, the interval is effectively zero. Thus, our concern was whether the facilitation effect would generalize to the shorter, more natural, interval.

A secondary question involved the interaction in Experiment 3. This suggested that the facilitation effect of a congruent adjective-noun pairing was somewhat stronger when the adjective was a real adjective than when it was a pseudoadjective. This outcome speaks against the modularity of inflectional processing and semantic

(i.e., stem) processing because it suggests the possibility that semantic characteristics of the adjective influence the speed of processing the noun's inflection. A second purpose of Experiment 4 was to determine if the interaction found in the previous experiment would still be found in the more natural interstimulus interval. If not, it could be claimed that the apparent nonindependence observed in Experiment 3 had just been an artifact of the abnormally long length of time between stimuli.

Method

The digitized stimuli created for Experiment 3 were used. The silent interval between the adjective and noun was reduced to zero for one set of stimuli and was increased to 800 ms for an otherwise identical set of stimuli. Two sets of 36 subjects each were run; nine subjects in each set were given one of the four permutations of adjective-noun combinations described in Experiment 3. One set of 36 subjects received the zero ms interstimulus interval stimuli (Short ISI) and a second set of 36 received the 800 ms interstimulus interval stimuli (Long ISI). None of the subjects had participated in Experiment 3. All procedures, including practice trials, were identical to Experiment 3.

Subjectively, a short ISI pair sounded like fast but normal speech; in contrast, a long ISI pair was perceived with a brief but distinct pause between adjective and noun, as if the speaker were enunciating carefully.

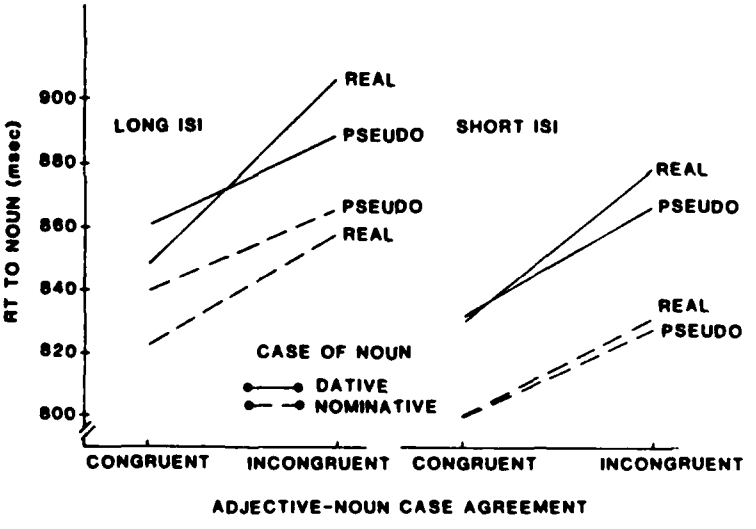


Figure 3. Lexical decision reaction time to real nouns preceded by real or pseudo adjectives with congruent or incongruent case matches and long (800 ms) or short (zero) interstimulus intervals between adjective and noun.

Results and Discussion

Figure 3 presents lexical decision RT means for real word noun targets only. Responses are presented as a function of ISI, the case of the noun (nominative or dative), and the grammatical congruency of the adjective-noun pairing.

Inspection of Figure 3 suggests three clear effects. Apparently, RTs were faster for short than for long ISIs. Also, responses to nouns inflected with the nominative case suffix appear to be faster than responses to dative case nouns. Finally, nouns in grammatically incongruent adjective-noun pairings are consistently responded to more slowly than in congruent pairings. Analyses of variance supported these

suggestions, to greater or lesser degrees. Least powerful was the effect of ISI; for the subjects analysis, $F(1,70) = 3.43$, MSe = 486, $p < .06$ and for the items analysis, $F(1,54) = 86.91$, MSe = 19, $p < .001$. For the Case of the noun, the subjects analysis gave $F(1,70) = 118.19$, MSe = 19, $p < .001$ and the items analysis gave $F(1,54) = 9.56$, MSe = 61, $p < .004$. Finally, for the main effect of Congruency, the subjects analysis gave $F(1,70) = 169.62$, MSe = 16, $p < .001$ and the stimulus analysis, $F(1,54) = 44.40$, MSe = 61, $p < .001$. In marked contrast to Experiment 3, there were no significant interactions for either ISI.

Most importantly, there were no significant effects involving the lexicality of the adjective; there is no evidence to suggest that the different effects of grammatically congruent and incongruent suffixes depended on whether the priming adjective was a real adjective or not. Only the congruency of the inflections made a difference, not the stem to which it was attached.

Despite the absence of any significant interaction between lexicality of the adjective and congruency, there is the hint of such a pattern for the long ISI condition and, more convincingly, a significant interaction in Experiment 3, in which the SOA was similar to the long ISI of the present experiment. We have no compelling explanation of these results except to suggest that, with long SOA, there is time for subjects to apply an experiment-specific strategy: because pseudoadjectives have only grammatical meaning, subjects may tend to drop them from short-term memory more readily and, for the long SOA, these pseudoadjectives (including, of course, their inflectional suffix) will often have dissappeared from memory before the appearance of the following noun target. Therefore, the observed congruency effect will be attenuated for nouns that follow pseudoadjectives, but only at SOAs that are long enough to allow substantial forgetting to occur.

GENERAL DISCUSSION

The present study offers evidence of essentially similar lexical decision processing for printed and spoken words. Further, and more importantly, it suggests that syntactic (inflectional) information and semantic information are initially processed by different, separable, mechanisms.

In Experiment 1, the recognition latencies for spoken nouns replicated the pattern obtained earlier by Lukatela et al. (1980) for printed nouns: fast reaction times to nominative case forms and slower, but equal, reaction times among the oblique case forms. Lukatela et al. had demonstrated that this pattern (the 'satellite' pattern) could not be accounted for by the individual case frequencies in Serbo-Croatian. Experiments 1 and 2 demonstrated further that the satellite pattern is not caused by differences in phonological structure among the different case forms of the same noun: The satellite pattern was unaffected by adjusting reaction times to the point within a word where the inflection becomes uniquely identifiable. Thus, inflectional (syntactic) differences alone appear to account for the satellite pattern. Confirmation that this syntactic effect is independent of word frequency was indicated by a finding of identical satellite patterns for high and low frequency nouns, even though high frequency nouns were faster overall.

In current linguistic theory, inflectional morphemes are described as having a distinct function and representation in the grammar. The results of the first two experiments suggest that this intuition has some counterpart in the cognitive mechanisms underlying word perception: that a separate device for processing inflection exists. One characteristic of the device is suggested by the fact that the case-dependent satellite pattern occurred even though, logically, a subject could have made a lexical decision after listening to the stem without waiting for or attending to the inflection (because the inflection was always correct and, therefore, redundant

with regard to the lexical decision). The fact that subjects, instead, were affected by the inflection suggests that the operation of the inflectional processor is mandatory during word recognition.

The suggestion of an autonomous inflection processor must be tempered, however, by the knowledge that the inflectional message (syntactic information about the case of the noun) cannot be interpreted solely on the basis of the inflectional suffix. As was discussed earlier, the phonological form of the suffix (e.g., -a, -e, or -u, etc.) cannot be assigned a case role unless the gender of the stem is known. This gender assignment is generally arbitrary and, therefore, must be represented lexically for each stem. Inflectional processing, then, must follow lexical access, an access that is based on the stem. This scenario of sequential events, in which lexical access for the stem precedes the activation of a separable inflectional processor, is, not surprisingly, in accord with the facts of Serbo-Croatian word formation, in which the stem always precedes the inflectional morpheme.

It is reasonable to ask whether this scenario is peculiar to the Serbo-Croation language or, instead, represents a universal tendency of languages for word recognition. One piece of evidence in favor of the latter hypothesis is the tendency for languages to use suffixes instead of prefixes for inflection (Greenberg, 1966). Some languages also make use of infixes or word-internal changes but only rarely does inflection precede the base. The temporal priority given to lexical access over inflectional processing may also be reflected in the developmental pattern of language acquisition; children produce violations of syntax even when they make no semantic violations. Moreover, the temporal precedence of lexical access over inflectional processing would follow as a plausible consequence of the evolution of the perception and production of syntax from an earlier, nonsyntactic, mind in the development of Homo Sapiens. Evolutionary changes that are extensions of earlier functions tend to build onto the previous system in a way that preserves the modular character of that earlier structure; the alternative would be to reform the entire preexisting system in order to accommodate the new function.

This view of the processing of inflectional information as being postlexical is consistent with the view of Seidenberg and his associates (e.g., Seidenberg, 1985; Seidenberg, Waters, Sanders, & Langer, 1984; Seidenberg, Tannenhaus, Leiman, & Bienkowski, 1982). They find that syntactic priming, as well as semantic (associative) priming, facilitates lexical decision but only the latter kind of priming, semantic priming, has a facilitating effect for naming. They interpret this to mean that a) lexical decision latency reflects both lexical and postlexical processing but naming latency reflects only lexical processing and b) syntactic processing is postlexical. Seidenberg et al. have compared lexical decision and naming only for the English language and only for printed material. Nevertheless, the similarity of interpretations for both English and Serbo-Croatian, two languages that implement grammatical meaning in different ways, inclines us to conjecture as a general principle that lexical processing occurs before inflectional processing.

One cautionary note: Katz and Feldman (1982) compared printed word recognition in English and Serbo-Croatian and found there were some differences between the two languages for the naming task although they appeared similar for lexical decision. For example, English naming was facilitated by semantic priming but Serbo-Croatian naming was not, a result attributed to the highly regular spelling-to-sound correspondence in the Serbo-Croatian orthography (Katz & Feldman, 1981). Thus, it is quite possible that naming in Serbo-Croatian may be accomplished differently than naming in English (although lexical decision may be similar) and, therefore, the techniques of Seidenberg et al. may not be appropriate for distinguishing pre- and postlexical processing in that language.

The third and fourth experiments further strengthened the evidence from the first two experiments that the inflectional processor a) is unaffected by semantic information, b) functions obligatorily in natural, rapid speech, and c) operates post-lexically. It was found that a noun was identified more easily when it was preceded by an inflectional suffix that was syntactically predictive. It did not matter whether the semantic content that accompanied the adjective was full (i.e., a real adjective stem) or was null (i.e., a pseudoadjective stem). Experiment 4 demonstrated that equal amounts of relative facilitation were produced by pseudo- and real adjectives. In addition, the relative facilitation effect was strong even when the interval between the adjective and the noun was as short as it is in normal rapid speech. The important finding of Experiment 4 was that the relative facilitation effect was not weaker when the adjective-noun interval was shorter. The effect did not dissapear when there was little or no time for conscious strategies to operate.

This suggests that the inflection processor is normally activated when listening to natural rapid speech and that its operation is not informed by the listener's expectations or biases.

Our results are consistent with a model of an inflectional processor whose major characteristics are (1) its inputs are limited to syntactic information (e.g., word class, gender, and inflectional suffix), (2) the process is not influenced by semantic information, (3) its operation, once initiated, is self-contained (non-interactive and, in particular, not informed by higher order cognitive processes), and (4) its output is specifically syntactic in nature (e.g., the meaning of a noun's case role and number). In Fodor's (1983) terminology, the model describes a system that is domain specific and informationally encapsulated. Our experimental evidence suggests that, in addition to these characteristics, inflectional processing is mandatory as well.

ACKNOWLEDGMENT

This research was supported by NIH grants HD-08495 to the University of Belgrade, Belgrade, Yugoslavia and HD-01994 to Haskins Laboratories. We are indebted to Aleksandar Kostić and Milan Savić for their advice and aid in all phases of the study. We gratefully acknowledge the assistance of Mira Peter in running subjects for Experiment 4 and Joan McCann for Experiment 3.

REFERENCES

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *The perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Cambridge, MA: MIT Press.

Fowler, C., Napps, S., & Feldman, L. B. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241-255.

Greenberg, J. H. (1966). *Universals of language*. Cambridge, MA: MIT Press.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*, 267-283.

Grujić, B. (1969). *Recnik englesko-serpskohrvatski sprpskohrvatski-engleski* (English-Serbocroatian Dictionary). (14th ed.). Belgrade, Yugoslavia: Prosveta.

Gurjanov, M., Lukatela, G., Moskovljević, J., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by inflected adjectives. *Cognition, 19*, 55-71.

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., & Turvey, M. T. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 692-701.

Katz, L., & Feldman, L. B. (1981). Linguistic coding in word recognition: Comparison between a deep and a shallow orthography. In A. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, NJ: Erlbaum.

Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 157-166.

Kempley, S. T., & Morton, J. (1982). The effect of irregularly related words in auditory word recognition. *British Journal of Psychology, 73*, 441-454.

Kostić, A. (1983) *Verb valence and lexical decision*. Doctoral Dissertation, University of Connecticut, Storrs, Connecticut.

Kostić, Dj. (1965). *Frequency of occurrence of words in Serbo-Croatian*. Unpublished manuscript, Institute of Experimental Phonetics and Speech Pathology, Belgrade, Yugoslavia.

Lukatela, G., Gligorijević, B., Kostić, A., & Turvey, M. T. (1980). Representation of inflected nouns in the internal lexicon. *Memory & Cognition, 8*, 336-344.

Lukatela, G., Kostić, A., Feldman, L., & Turvey, M. (1983). Grammatical priming of inflected nouns. *Memory & Cognition, 11*, 59-63.

Lukatela, G., Moraca, J., Stojnov, D., Savić, M., Katz, L., & Turvey, M. T. (1982). Grammatical priming effects between pronouns and inflected verb forms. *Psychological Research, 44*, 297-311.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhuis (Ed.), *Attention and performance X: Control of language processes*. Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M., & Tannenhaus, M. K. (in press). Modularity and lexical access. In I. Gopnik (Ed.), *McGill studies in the cognitive sciences*. Norwood, NJ: Ablex.

Seidenberg, M. S., Tannenhaus, M. K., Leiman, J. L., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition, 12*, 315-328.

Smith, S., Katz, L., & Macaruso, P. (1984, April). *Lexical representation of inflected verb forms*. Paper presented at the Eastern Psychological Association Meeting, Baltimore, MD.

FOOTNOTES

*Cognition, 25, 1987, 235-263.

**Also University of Connecticut

†Also Yale University

††University of Belgrade

Talkers' Signaling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction*

Carol A. Fowler† and Jonathan Housum††

An experiment examines talkers' utterances of words produced for the first time in a monologue ("new" words) or for the second time ("old" words). The finding is that talkers distinguish old words by shortening them. Two experiments show that old words are less intelligible than new words presented in isolation, but probably are not less identifiable in context. We infer that talkers may attenuate their productions of words when they can do so without sacrificing communicative efficacy. Old words can be reduced because they are repetitions of earlier presented items and because of the contextual support they receive. Two final experiments show that listeners can identify new and old words as such and that they can use information that a word is old more-or-less as they would use an anaphor to promote retrieval of the earlier production in its context.

Bolinger (1963, 1981) suggests that when talkers utter words that are unusual in their contexts, they lengthen them. In his example, speaking of the return trip of a person who had ridden his lawn-mower cross-country, one might say: "he mowed home." In that context, according to Bolinger, *mowed* is lengthened as compared to its duration in a sentence discussing the more usual uses of lawn-mowers. Possibly, then, talkers lengthen words that have little contextual support, or more generally, that have little or no other information than their acoustic signal to specify their identity.

Perhaps compatibly, Lieberman (1963) found a difference in intelligibility of redundant and nonredundant words presented in noise. He found that a word, for example, *nine*, that had been produced in an uninformative context ("The word that you will hear is _____") was more intelligible excised from the sentence and presented in noise than the same word originally produced in a more informative context ("A stitch in time saves _____"). Hunnicutt (1985) has partially replicated and has extended these findings.

An inference from this set of observations and findings taken together is that talkers aim to provide an acoustic signal for a word that is sufficiently informative for listeners to identify the word. If the word is probable in its context, talkers may provide a reduced, acoustically less informative version of the word than if the word has a low probability or is not redundant.

Why might a talker vary his or her production of a word in this way? Two mutually compatible reasons may be offered. One is that the reduced versions of words require less articulatory work to produce, and talkers may choose to do less work when they can get away with it without sacrificing communicative efficacy (cf. Koopmans-Van Beinum, 1980). An entirely different reason is suggested by extension of Chafe's

theorizing. Chafe (1974) proposes that talkers provide information to listeners in the way that they produce words to help them distinguish "given" and "new" information in discourse.

Given information is information shared by talker and listener; but, more than that according to Chafe, it is information that the talker presumes is currently foregrounded in the listener's awareness--because it has just been mentioned or because the listener is currently looking at the thing to be named, etc. By reducing their productions of words reflecting given information, talkers thereby highlight "new" information and draw the listener's attention to it. In this theory, then, the talker deploys reductions in a systematic way to highlight the most informative words in an utterance.

Chafe's "given" information is not the same as Lieberman's or Bolinger's high probability words. Nor is the reduction he writes of necessarily the complement of the augmentation noticed by Bolinger. Whereas Chafe writes of talkers lowering their voice pitch and destressing words conveying given information, Bolinger writes of a durational lengthening of low probability words. Nonetheless, there is enough family resemblance across these sets of observations and findings to warrant asking whether they may not point to some interesting hypotheses concerning the talker's deployment of lengthening or reduction in speech and its consequences for the listener. Possibly, talkers attenuate their productions of a word when they can without sacrificing the word's identifiability; in order not to sacrifice identifiability, they can reduce only words whose identity is determined in part by other information available to the listener. (We will call such words, words that provide "old" information.) If this scenario is accurate, then, the talker's deployment of reductions and of more careful productions is systematic, and they can provide information to a listener that the concept named by the reduced (or augmented) word is "old" (or "new").

The experiments reported here are designed to test these hypotheses in a preliminary way.[1] They are not designed to test Chafe's or Bolinger's proposals directly, but rather to address the more general hypotheses that the foregoing summary of the literature suggests.

EXPERIMENT 1

Neither Bolinger nor Chafe provides measurements of talkers' productions of low probability or given words. One reason why they disagree on the acoustic manifestations of augmentation or attenuation, then, may be simply that they noticed different aspects of the acoustic consequences of reduction and augmentation. The first experiment is designed to measure talkers' productions of new and old words in speech.

For these preliminary investigations, we decided to use spontaneous speech produced by talkers in a natural, or at least a nonlaboratory, setting. This has the advantages over speech collected in the context of a controlled experiment that talkers really are attempting to communicate something to someone and that they are unaware that the way in which they are speaking will be of interest to an experimenter. The procedure has disadvantages too. One is that the investigators have no control over the talker's use of new and old information; they must make use of whatever is said. A more serious problem is that the contexts in which a particular word appears as new or as old information are different. This is problematic because talkers use duration, voice pitch, and amplitude for multiple purposes, not just as indices of old- or new-ness. In spontaneous speech, therefore, there will be other uses of these variables that will serve as sources of random noise in the measurements.

Accordingly, whereas we can be confident that talkers do use a variable systematically if we find consistent differences in its values on new and old words, we cannot be confident that talkers do not use a variable just because we find no significant effects of it in our data.

For the purposes of Experiment 1, we defined a "new" word as one produced for the first time in a passage and an "old" word as a repetition of a word spoken once before in the passage, however far back. We looked only at first and second productions in the experiment and asked whether second productions of words are shorter and lower in the fundamental frequency and amplitude of their stressed vowels than first productions. Obviously, this operational definition of "old" and "new" does not provide an entirely valid indicator of redundancy, given-ness, or high and low probability. That is, a word may be old because a synonym for it has been presented earlier; too, many new nouns are replaced by pronouns when they are old. However, in the passages we used there were many examples that fit our definition. The definition has the advantage of allowing us to look at productions of different tokens of the same word when they are new and old.

Methods

Materials. The major source of evidence for this experiment is a monologue from Garrison Keillor's radio program, "*A Prairie Home Companion.*" The monologue, titled "Sylvester Krueger's Desk," lasts 18.5 minutes and purports to describe Keillor's days as a fourth grader in school in the imaginary town, Lake Woebegon (Gospel Birds, cassette tape, 1985). Although the monologue is not extemporaneous, as most conversation is, it was not read, and the speech sounds spontaneous and natural.

The passage was transcribed, and 35 pairs of words were selected for analysis. Criteria for selection were that a word occur at least twice in the passage and that, if relevant, it refer to the same object or event on both productions (so "match" referring to a "tennis match" on one occasion and as a way to light a fire on the other would be excluded). If a word occurred more than twice, just the first and second occurrences were used. Words that are chronically highly probable ("*of*," "*the*") were excluded. Also excluded were pairs of words in which one production was finally lengthened (usually because it occurred at the end of a major syntactic boundary; Cooper & Paccia-Cooper, 1980) but the other was not. Otherwise words (including some names and some phrases, such as "Labor Day" and "ten dollar bill") were considered eligible for selection, and most eligible pairs were selected. First productions were positioned nine words from the beginning of a sentence on average and 12 words from the end; second productions were the reverse: 12 words from the beginning of a sentence on average and 9 words from the end. In an analysis of variance, the interaction between first or second production and distance from the beginning or end of a sentence was marginally significant, $F(1,34) = 3.51$, $p = .07$. However, neither new nor old words tended to fall very close to either sentence beginnings or ends. A sample paragraph from the monologue with selected items underlined appears in Appendix A.

Five additional samples of speech were taken from interviews broadcast on the MacNeil-Lehrer Newshour and videotaped by a colleague for another purpose. They included separate interviews with two congressmen, two senators, and one newsperson. The shortest of these five passages contained just nine eligible word pairs. All of these were selected; in the other passages, the first nine eligible pairs were selected.

Procedure. Selected words were filtered at 10 kHz, sampled at 20 kHz, digitized, and stored on the hard disk of a computer (New England Digital Company). Three measurements were made of each word: the word's duration, the average

fundamental frequency (F_0) of its lexically-stressed vowel, and the peak amplitude of the same vowel.

All measurements were made from a waveform display. Duration measurements were made using visual and auditory evidence of word onset and offset. Zero crossings were identified in the waveform at locations where the word looked and sounded as if it started and ended. Measurements of F_0 and amplitude were confined tc the lexically-stressed vowel. (For some items, for example "Labor Day," there is more than one lexically-stressed vowel; in those cases, we measured the phrasally more prominent of the two stressed vowels [in the example, /ey/ from "*Labor*"].) F_0 measurements were obtained by counting pitch pulses in the selected vowel, measuring the duration they spanned, and transforming the measures to Hz values. Amplitude measures were taken from the pitch pulse in the stressed vowel with the highest amplitude; measures were in volts.

Measurements were made by the first author, but a sample of them was checked by a research assistant naive to the purposes of the experiment.[2] The sample included 20 of the 70 selected words from the Keillor passage. The 20 included 10 new words and 10 old words. These were selected randomly with the constraint that the new and old words be chosen from different pairs. Correlations between the two sets of measurements on the 20 words were .99 for duration, .94 for F_0, and .91 for amplitude.

Results and Discussion

Table 1 presents the findings on the 35 pairs of words from the Keillor monologue and below that, the five sets of nine words from the remaining passages. All comparisons reveal mean differences in the predicted direction if old words were attenuated as compared to new words. In a MANOVA with new/old as an independent variable and (log transformed) duration, amplitude and fundamental frequency as dependent measures, the effect of the independent variable was significant, $F(3,32) = 3.82$, $p = .02$. In univariate tests, the effect of duration, $F(1,34) = 9.83$, $p = .004$, and amplitude, $F(3,32) = 4.42$, $p = .04$, were significant; the effect of fundamental frequency was marginal, $F(1,34) = 3.20$, $p = .08$. A MANOVA was performed on the data from all six talkers, with talker and new/old as independent variables, and the effect of the new/old variable was significant once again, $F(3,46) = 3.27$, $p = .03$. However, in this instance, only the effect of duration was significant in univariate tests, $F(1,48) = 9.28$, $p = .004$. All talkers had overall shorter old than new words; four of six had lower amplitude old than new words; just two of six had lower frequency old words. In the same analysis, there was a significant effect of talker on the dependent measures; however, the interaction between talker and the new/old variable did not approach significance.

TABLE 1

Measurements of old and new words from the Keillor Passage and the five other passages in Experiment 1. (Measurements are in Ms, Hz, and Volts, respectively.)

	DURATION		F_0		AMPLITUDE	
	New	Old	New	Old	New	Old
Keillor	562	492	119	110	1.12	1.03
Others	436	395	135	134	1.92	1.77

In the analysis just reviewed, duration, but not F_0 or amplitude showed reliable differences depending on whether a word was being used for the first or second time.

However, effects even on duration were not perfectly consistent. In the Keillor passage, 25 of the 35 words (71%) had shorter "old" than "new" words; in the remaining pairs, the direction of difference was reversed. Moreover, when shortening was observed, it varied substantially in amount from 4 ms to 414 ms. Some of this inconsistency and variability can be ascribed to the fact that the speech was spontaneous and, therefore, many sources of variability in duration were uncontrolled. However, possibly in addition, shortening may differ in amount according to some variables relevant to the old/new dimension.

One source of variability in shortening is the duration of the word when it is produced as new. Possibly, longer words generally have more room to shorten and so they may shorten more. This was the case in the Keillor monologue ($r = .46$, $p < .01$). A more interesting source of variation is the distance between the repetitions of a word. That is, talkers may feel free to shorten their productions of words that have just been said, but not words so far back in the conversation that listeners may not remember them (cf. Chafe, 1974). Among the 35 word pairs in the Keillor monologue, the second production followed the first by four words at the shortest lag and by 512 at the longest. The correlation between distance (in number of words) and shortening was exactly zero; with effects of the duration of the first production partialed out, it was .13, a nonsignificant difference in the wrong direction for the hypothesis.[3]

A final source of variability in shortening was sought in the topicality of the word pairs. Chafe (1974) proposes that talkers attenuate their productions of a word if they believe that the concept named by the word is already at the focus of listener's attention. Presumably, this would include words central to the topic of the discourse, but not the less topical words. Accordingly, we asked whether shortening would correlate positively with judged centrality of a word's meaning to the topic of its sentence or of the monologue itself.

We obtained topicality ratings in a subsidiary experiment, the methods of which are described in Appendix B. In that experiment, 10 subjects read the transcription of the Keillor monologue through and then filled out a rating sheet. On the sheet, the 35 word pairs were listed along with the page and line number in the transcription where each critical word occurred. Subjects were asked to give three ratings for each pair. They were to use a 10-point scale to rate the importance of the meaning of the word to the topic of the monologue as a whole and to rate the importance of each token of the word to the topic of its own sentence.

Only the first rating predicted shortening significantly, and that correlation was negative, contrary to prediction ($r = -.38$, $p = .02$). That is, words judged most important to the topic of the monologue were shortened less than less important words. Neither this correlation nor the correlation with distance is consistent with an idea that talkers only shorten words they consider to be currently at the focus of the listener's attention. Instead, the correlation with topicality suggests that talkers are least willing to shorten the most important words of the passage.

In summary, our findings so far indicate that talkers do attenuate their production of many "old" as compared to "new" words in discourse, the attenuation appears to take the form largely of shortening, and the shortening is least for words most central to the topic of the conversation.

We might ask why a talker would shorten old words. One answer that is likely to be correct is that attenuated productions, like casual speech more generally (Koopmans-Van Beinum, 1980; Zwicky, 1972), is easier to produce than slower, more formal productions. We will not pursue this hypothesis here. Instead, in the next two experiments, we assume that, in some sense, the talker wants to attenuate productions where possible, and we ask what allows him or her to do so.

EXPERIMENT 2

In the present experiment and the next, we consider three possible conditions that may allow talkers to attenuate their productions of words. One is that the reductions may be so slight as to leave intelligibility of the words unimpaired. This hypothesis is unlikely to be correct; if it were, then talkers presumably would attenuate their productions even of new words. A second possibility is that talkers attenuate words that have been produced before, because, in identifying a repeated word, listeners can benefit from having heard it once before in the discourse. This benefit may have two, possibly related, sources. One source is a repetition priming advantage in identification of or lexical decision to previously presented words (e.g., Kempley & Morton, 1982; Fowler, Napps, & Feldman, 1985). A second source is simply that once having figured out a word's identity, especially if it is unfamiliar (for example, a name, such as *Sylvester Krueger* in the monologue), a listener need not figure it out again based only on the acoustic signal; he or she can use the signal as a way of retrieving the word from memory. A final reason why talkers may be able to attenuate their productions of some words in a passage is that the words may be partially specified by their context. Possibly, the second productions of words are, on average, more redundant with their context than are first productions. Experiment 2 tests the first two possibilities; Experiment 3 tests the last.

Method

Subjects. Subjects were 36 students at Dartmouth College who participated for course credit. They were native speakers of English who reported normal hearing.

Materials. Two versions of a test audio tape were created; each consisted of the 35 word pairs from the Keiller monologue measured in Experiment 1. The words were excised from the monologue (using the zero crossings identified in Experiment 1 as word boundaries) and were presented at a rate of one every 5 seconds. Test orders on both tapes consisted of two blocks of 35 words. One member of each of the 35 word pairs occurred once in each block. In one block, 17 items were first productions and 18 were repetitions; the other block had 17 second productions and 18 first productions. Words were differently randomized in each block. The two test tapes were complements of one another. That is, where the first tape had the first production of "antique" as its 30th trial, the second tape had the second production of "antique" in that same slot. In this way, both productions of the words of every pair appeared equally often in the first and second blocks of the tape. Eighteen subjects listened to each tape.

Procedure and design. Subjects were run in groups of 2 to 3. They were told that they would be listening over headphones to words, names, or short phrases excised from a monologue. Their task was to identify each item if possible by writing it on the answer sheet; or if they could not identify an item, to write down as much of it as they could identify. In addition, they were to circle a number from 1-5 on their answer sheet expressing their confidence in their answer. A rating of 5 represented the highest degree of confidence and 1 the lowest. There was one independent variable, word history (new, old); the dependent variable was accuracy.

Results and Discussion

Subjects' responses were scored in two ways. First, answers were scored correct only if they were completely correct. (So, for example, the answer "plum" to the word "plump" received no credit for its close approximation to the target word.) In a second scoring method, answers were given scores representing the proportion of the phonemes in the stimulus string that were represented in the correct serial order in the response string. Because this scoring procedure gave exactly the same outcome in pattern and statistically as the first, we will not describe it further.

Results are given in Table 2. Subjects made more errors on old words than new words, $F(1,35) = 10.31$, $p = .003$, and more errors on words presented in the experiment for the first time than words presented for the second time, $F(1,35) = 7.10$, $p = .011$. The interaction of the variables was not significant, $F < 1$. Neither independent variable had significant effects in an analysis using items as the random factor. However, one reason for this outcome was a ceiling on performance on many items. In the condition associated with the lowest performance (that is, old words that appeared in the first block of trials), subjects achieved perfect accuracy on over half of the words (18 of 35). Of words on which some errors were made, the majority of errors in both blocks were made on repetitions (60%), and the majority of errors were made on words in the first block (72%).

TABLE 2

Percent errors and confidence judgments on new and old words and on the first and second blocks of the isolated-words perception test of Experiment 2.

	Occurrence in Monologue	
Block of test	First	Second
	Percentage Errors	
1	11.6	16.2
2	8.8	12.4
	Confidence Judgments	
1	4.47	4.26
2	4.69	4.49

There is, of course, the possibility that the improvement subjects show on the second block of trials is due to a more general practice effect than the one we have been considering. That is, subjects' ability to identify words excised from context may improve with experience, and the improvement on members of word pairs that are presented second as compared to first may be a consequence of their later presentation in the test list. We looked for evidence of a practice effect of this sort by comparing performance across trials 1-12, 13-24, and 25-35 in block 1. Contrary to expectation, if the improvement for block 2 items was a general practice effect rather than a specific effect of having heard other tokens of block 2 words before, performance was nonmonotonic over the successive thirds of the first block. Performance was lowest in the middle third and slightly better in the two flanking thirds; performance in the first and last third was nearly identical.

Analysis of confidence judgments gave an outcome similar to the analysis of error percentages. In both the subjects and items analyses, effects of the old/new variable (subjects: $F(1,35) = 66.66$, $p < .001$; items: $F(1,34) = 5.47$, $p = .02$) and of block (subjects: $F(1,35) = 32.08$, $p < .001$; items: $F(1,34) = 6.64$, $p = .01$) were both significant and consistent with the outcome on accuracy. The interaction was nonsignificant in both analyses.

Errors on old words in Experiment 2 correlated significantly with the duration difference between new and old words found in Experiment 1 ($r = .36$, $p < .05$). That is, words that had been shortened substantially in their second production were less intelligible, excised from context and presented in isolation, than words that had been shortened less. Likewise, the more an old word had been shortened, the bigger the difference in intelligibility between the new and old word in Experiment 2 ($r = .42$, $p = .01$). Finally, the more a word had been shortened, the greater its gain in intelligibility when it was presented in the second as compared to the first block of

trials (that is, when it was preceded by another production of the same word in the first block; $r = .32$, $p = .05$).

Discussion

Earlier, we proposed three possible answers to the question of what allows a talker to attenuate his or her production of an old as compared to a new word. Two of these proposed answers were addressed in the present experiment. One was that the attenuation was insufficient to affect intelligibility of the word based solely on its own acoustic signal. This was disconfirmed in the present experiment; old words were less intelligible than first productions.

The second answer was that a listener may be able to identify a word better if it follows an earlier token of the same word. This was the case in Experiment 2. Words in the second block of the experiment were identified more accurately than words in the first block. Indeed, the gain in intelligibility accruing to words in the second block was nearly enough to offset the loss of intelligibility owing to the reduction factor. That is, new words in the first block of trials were associated with an error rate of 11.6%. Old words in the second block had an error rate of 12.4%--a small difference only slightly in favor of the more careful productions. This leaves just a little work for the effects of context, examined in Experiment 3, to do for the attenuated words.

Before turning to that experiment, we should comment on a different aspect of the outcome of Experiment 2, not directly related to the questions under study. It is that words excised from context were highly intelligible in this experiment. Performance averaged about 88% correct and confidence was very high. Moreover, the performance measure almost certainly underestimates the intelligibility of words based on their own acoustic signals, but still presented in the context of the discourse. That is, the F_0 pattern, the amplitude contour, and the duration of the word reflect, in part, the word's position in its sentence and most probably its role in the sentence as well. These patterns will be at best uninterpretable when the word is presented excised from its context; at worst, they provide misleading information in their new setting. Any coarticulatory influences from neighbors likewise will present misleading information in an excised word. On the other side, Keillor was the slowest of the six talkers studied in Experiment 1 and he was speaking to a large audience so that the high intelligibility of his speech may overestimate that of talkers in conversation for example (see, e.g., Pickett & Pollack, 1963).

In Experiment 3, we consider the role that context may play in facilitating identifiability of a target word.

EXPERIMENT 3

Listeners to a repetition of a word have another advantage besides having heard the word produced once before. By the time the repetition occurs, they are farther into the discourse and so they may have more information about the topic; possibly, therefore, the second occurrence of a word may generally be more redundant with its context than is the first occurrence with its context. In this experiment, we estimate that possible difference in redundancy by asking subjects to guess the target words of Experiments 1 and 2 in their contexts.

Method

Subjects. Subjects were 14 students at Dartmouth College who took part in the experiment for course credit. They were native speakers of English.

Materials. Two versions of the transcribed monologue, "Sylvester Krueger's Desk," were prepared. In each version, one member of each of the 35 word pairs from Experiments 1 and 2 was selected to serve as a test word. In one version, there were 17

new items and 18 old items. The other version was the complement of the first with 17 old items and 18 new ones. Seven subjects received each version of the monologue. The passages were printed on a computer terminal and subjects made their guesses by typing a word or words into the computer.

Procedure. Subjects were run individually. They sat in front of a computer terminal on the screen of which the monologue was gradually printed. After printing a full screen of text, the program waited for input from the subject before scrolling upward and adding more text. Thirty-five times during presentation of the text, a question mark appeared on the screen and the program stopped printing. Subjects were instructed to read the text as it appeared on the screen. When they saw a question mark, they were to try to guess the next word, name, or short phrase. They made their guesses by typing them on the terminal keyboard and hitting the return key when they were finished. The program then continued printing the text, taking up where it had left off (and therefore, providing subjects with feedback concerning their guess). Subjects were told that they could guess just one word or more than one as they wished.

Answers were scored correct if the first word typed by the subject matched the first word of the passage after the point at which the question mark had appeared. The session lasted about one-half hour.

Design. The experiment had one independent variable, whether the guessed words were old or new. Subjects were crossed with the independent variable. The dependent variable was accuracy measured as the percentage of words guessed correctly.

Results and Discussion

On average, subjects guessed 18.3% of the new items correctly and 31.1% of the old items. This was a significant difference in the analysis by subjects, $t(13) = 3.79$, $p = .002$, with 13 of the 14 subjects showing effects in the predicted direction. The analysis was not significant by items, however: $t(34) = 1.54$, $p = .13$. The items analysis was nonsignificant because of a floor on performance on many items. That is, subjects made no correct guesses on either occurrence of over one-third of the test items. Of the 23 items on which at least one correct guess was made, 13 old items showed better performance than their corresponding new items, five new items were superior to their counterpart old items, and five pairs showed no difference at all.

Two aspects of this outcome are interesting. One is that there is a tendency for old items to be more predictable from their contexts than are new items from theirs. Perhaps more notable, however, is the finding that subjects do not very often succeed in guessing the exact next word from context. Our finding that, on average, subjects guess correctly 24.7% of the time is nearly identical to a finding by Gough, Alford, and Holley-Wilcox (1981) using a procedure in which subjects are asked to guess each successive word of a passage. This is not to say that subjects cannot often guess the *content* of the forthcoming word. Indeed, their guesses often were very close in content to that of the forthcoming word (e.g., "old" for "antique" and "homerun" for "double"). However, the guesses infrequently corresponded to the exact next word in the passage. As Gough et al. (1980) conclude, guessing from context is unlikely to play a role in ordinary reading or listening.

However, in conjunction with the rather good information for each word's identity that Experiment 2 suggests that the talker provided, context can help to eliminate mishearings. (For example, it can distinguish "plum" from "plump.") Moreover, the present experiment suggests that the contextual support does tend to be better for the old words, which may require more support, than for the new items.

Returning to the question posed earlier as to what allows a talker to produce reduced versions of old words, then, we can suggest that listeners can recoup the consequent decrement in intelligibility in two ways. They can benefit from having

heard the word once before and they can benefit from the more constraining context in which the old word tends to occur.

In the final experiments, we ask whether the reduction of old words not only fails to impair the intelligibility of the talker's message, but, in addition, may even provide useful information to the listener.

EXPERIMENT 4

Here we ask whether listeners can tell whether a particular utterance of a word is new or old. If they can, then possibly they can use information that a word is old to distinguish given from new information as Chafe proposes, or they can use it as a listener uses an anaphor to recover prior mention of the concept in its context (e.g., McKoon & Ratcliff, 1980).

The design of the present experiment presents listeners with a more difficult judgment than they confront in listening to continuous discourse. We presented new and old words in isolation and asked subjects to identify each as new or old. As we have pointed out before, the fundamental-frequency contour, amplitude contour, and durations of a word in part reflects its position and role in a sentence. Pulled from context, these acoustic properties of words may be more than uninformative; they may be misleading. However, if, even under these adverse conditions, listeners can make the distinction, we can be sure that they can make it in context too.

Method

Subjects. Subjects were 18 students at Dartmouth College who received course credit for their participation. They were native speakers of English who reported normal hearing.

Materials. We used the audio tapes created for the identification test of Experiment 2.

Procedure. Subjects were run in groups of 1, 2, or 3. The experimenter explained to them that talkers attenuate their productions of words the second time they say them as compared to the first time and that the purpose of the experiment was to learn whether listeners could tell, from the way a word is spoken, whether it is being said for the first or the second time. The subjects' task was to listen to each word on the tape and to write either a 1 or a 2 on their answer sheet, 1 signifying a guess that the talker had not uttered the word before, 2 signifying a guess that the talker was saying the word for the second time. We also gave subjects the information that the tape consisted of 35 pairs of words, each pair consisting of a first and a second production of a word; therefore, on average, they should distribute their responses evenly among 1's and 2's. Nine subjects listened to one version of the tape and the remainder listened to the second version.

Results

Over the 70 trials of the experiment, subjects averaged 60% correct. Fifteen of the 18 subjects performed numerically better than chance, two subjects were at chance, and one was numerically below chance with 33 of 70 items correct. A paired t test comparing performance to the chance value of 50% was highly significant, $t(17) = 6.02$, $p = < .001$.

Although subjects found the task very difficult and made many errors, almost all of them could do the task. This lends some encouragement to an idea that listeners do have information available in the way words are pronounced in spoken discourse that indicates whether the word has been uttered before. In the final experiment of this study, we ask whether listeners use that information in comprehending speech.

EXPERIMENT 5

In this experiment, we ask whether reductions of words when they are old actually promote comprehension by facilitating integration of related material in the discourse. That is, attenuated old words provide information in the way they are said that a word--and presumably what is being said about it--refers back to something said earlier. In this way, reduction may serve a role similar to the role of pronouns and other anaphors. A pronoun is generally less audible than the word, name, or phrase it replaces; it is short and often destressed. Despite that, it may be more informative than an exact repetition of the item it replaces because, being a pronoun, it announces its referent as having been mentioned before. This may facilitate the listener's retrieving the previous relevant information and connecting it with what is being said now. Other anaphors, for example, referring to a previously mentioned car as "the vehicle" may likewise signal that the label names a previously mentioned concept.

Experiment 5 was designed to ask whether reductions work in the same way. The experiment was designed by analogy with work done on anaphoric reference by McKoon and Ratcliff (1980). In two experiments, they showed that anaphors activate not only the words they replace, but also other words from the same proposition as the words they replace.

A sample paragraph from their experiments follows: "A burglar surveyed the garage set back from the street. Several milk bottles were piled at the curb. The banker and her husband were on vacation. The burglar/the criminal/a cat/ slipped away from the streetlamp." In the experiment, different groups of subjects saw the three versions of the last sentence. In one version, a word from the first sentence reappeared in the last sentence; in another, an anaphor appeared, and in a third, an unrelated word appeared. After reading a paragraph, subjects were given a test word on which to make an "old" or "new" response depending on whether the word had appeared or not in the preceding paragraph.

McKoon and Ratcliff found that anaphors (e.g., "the criminal") in the last sentence of the paragraph were as effective as repetitions of previously mentioned words (e.g., "the burglar") in activating the test word (e.g., "garage") from the first sentence of the paragraph; both versions of the last sentence led to faster "old" decisions to the test word than did the version with an unrelated subject noun phrase ("a cat").

In a second paradigm, subjects read two paragraphs and then made old/new decisions to a list of test words. In that experiment, a word from the last sentence (e.g., "streetlamp") was judged more rapidly preceded by "burglar" than preceded by a word from the other paragraph. This outcome occurred both for subjects who had seen "burglar" in the last sentence and for subjects who had seen "criminal."

Although priming differences between the original word and the anaphor were very small and nonsignificant, numerical differences favored the anaphor in both experiments. In the present experiment, we used a procedure similar to those of McKoon and Ratcliff to ask whether a reduced version of a spoken word might serve as a better reminder of words in a sentence containing a nonreduced version of the same word than would the nonreduced version itself.

Method

Subjects. Subjects were 33 students at Dartmouth College who received course credit for their participation. They were native speakers of English who reported normal hearing. Data from two subjects were eliminated because of poor performance (near chance accuracy in one instance, response times averaging twice those of the remaining subjects in the other).

Materials. Once again, the Keillor monologue was used. From the monologue, 42 prime-target pairs were selected on which subjects would make judgments whether or not the word had occurred before in the passage. (These were called "old"/"new" decisions in speaking to subjects; however, to avoid ambiguity with another use of "old" and "new" in this manuscript, we will call them "yes"/"no" recognition decisions.) At 42 selected locations in a re-recording of the monologue, a 1000 Hz tone was placed on the other channel of the tape than the channel used to present the monologue to subjects. This tone, input to a computer, caused the program running the experiment to stop the tape recorder and to present a warning tone followed by two words on which subjects made speeded recognition judgments. We will call the first word of each pair the "prime" and the second word the "target" for that trial. Response times were measured from prime and target onset.

Of the 42 prime-target pairs presented to subjects, 14 were critical pairs and the remainder were fillers. In the critical prime-target pairs, both words had occurred recently (12 syllables back on average) in the monologue and so the correct recognition judgment was "yes." In all of the critical pairs, the prime was one of the 70 words measured in Experiment 1 and further studied in Experiments 2-4. In seven pairs, the prime was the same version of the word produced recently in the passage and it was the first occurrence of the word in the monologue. In seven pairs, the prime was not the same version of the word as that just produced in the monologue and it was the old version of the word. On these trials, the target was some other word, near the nonreduced prime and in the same sentence as the nonreduced prime in the monologue. The 14 critical primes were selected based on their distribution throughout the monologue; they were not selected based on the durational difference between new and old versions of the primes. Table 3 shows how a trial was organized in the experiment.

TABLE 3

Sample Trial from Experiment 5.

Oh that smell about it, same blackboard, same
portraits of Washington and Lincoln up front and center

↓

Warning tone (100 ms)

↓

SOA = 1100 ms

↓

$blackboard_1$ (Group 1)/$blackboard_2$ (Group 2)

↓

SOA = 1500 ms

↓

portraits

Two versions of the experiment were run, one on 16 subjects and the other on the remaining 15. The versions were complementary so that if subjects in the first group had the new version of a word as a critical prime, subjects in the second group had the old version on the same trial.

Filler trials included six trials in which both prime and target had not occurred in the monologue at the point where they were tested. (That is, the correct response to both prime and target was "no.") In this and other filler trials, words on which a "no" response was correct were selected from the monologue but from a location farther on

than the point where the words were tested. Eleven trials each were "yes"-"no" and "no"-"yes" trials. Trials of the various types occurred in quasi random order and occurred at irregular intervals throughout the monologue. The first critical trial was the fourth trial of the experiment.

Design. There was one independent variable, whether the critical prime was a first or a second production. Dependent variables were response time to the target and accuracy of response to the target. In addition, we looked at response times and accuracy to primes depending on whether they were first or second productions.

Results and Discussion

Response times to targets were included in the analysis only if the response was accurate and if the response to the prime had been accurate. There were no correct response times to critical primes or targets slower than 2500 ms; no responses were deleted from the analysis because of their duration.

Table 4 presents the mean response times and proportions correct for critical primes and targets. On targets, the accuracy measure reflects the number of correct responses independent of accuracy on primes.

TABLE 4

Average response times and proportions of correct responses to new and old primes and to targets preceded by new and old primes. Data from Experiment 5.

	PRIME		TARGET	
	New Prime	Old Prime	New Prime	Old Prime
RT	834	793	758	719
Accuracy	.92	.90	.89	.91

Responses to reduced primes were overall faster than to first productions. The difference was significant in the subjects analysis only, $F(1,29) = 11.31$, $p = .002$. The same analysis showed no effect of group and no interaction of group by prime type. The difference between new and old primes was not significant in an analysis by items, $t(13) = 1.32$, $p = .21$. Of the 14 critical items, nine showed a difference favoring the reduced prime.

The difference in response times to the new and old primes, significant in the analysis by subjects may, in any case, reflect only the duration difference between reduced and unreduced words. This difference averaged 89 ms for the 14 critical items of the experiment.

Response times to targets are faster following old primes than following new primes. Results are weak but significant in both subjects and items analyses (subjects: $F(1,29) = 4.15$, $p = .048$; items: $t(13) = 2.25$, $p = .04$). The small accuracy difference also favors targets preceded by old primes; however, the difference did not approach significance in either analysis by subjects or by items.

The significant difference in reaction time apparently cannot be explained simply as faster response times to targets that follow short primes or that follow fast responses to primes. Correlations between response times to targets and prime durations, and between response times to targets and response times to primes (computed separately on new and old primes to eliminate effects of the independent variable) are uniformly nonsignificant.

In conjunction with Experiment 4, the present experiment shows both that listeners can distinguish reduced from unreduced versions of a word and that they

can use the perceived reduction as information that a word has been mentioned before to facilitate recall of the word's prior context.

GENERAL DISCUSSION

We have found that talkers attenuate their productions of old words and that the identifiability of these redundant words is affected if the words are presented in isolation, but is probably not affected for the words in context. Finally, we have found that listeners can identify words as old or new, and they can use information that a word is old to facilitate integration of related material in a discourse.

If talkers reduce old words, as we suppose, for "selfish" reasons--an idea, it is true, that requires experimental test--then the present study reveals an interesting example of a sort of "symbiotic" relationship between talkers and listeners.

Talkers may reduce their productions of old words because it is easier to produce reduced than careful versions of words, and because listeners do not need as good a signal for an old as for a new word. Listeners do not need as good a signal because, having heard a word before, they find it relatively easy to identify it a second time, and because the context of an old word tends to be more constraining than that of a new word. By reducing old as compared to new words, however, talkers deploy reduction systematically and therefore, reduction (or on the other side, careful articulations) can provide information to a listener that a word relates back to something said earlier (or does not). As Experiment 4 shows, listeners can tell reduced from unreduced words even under quite adverse conditions in which the words are excised from their context. Experiment 5 shows that they can take advantage of the information provided by reductions to retrieve the earlier context of the word.

Possibly, this instance of a behavioral systematicity that is beneficial for different reasons both to talkers and to listeners is not unique to production and perception of new and old words in speech. Indeed, possibly this confluence of mutual benefits may promote the perpetuation of various systematic behaviors in a language and across languages.

That is, there may be other examples in which talkers produce speech in certain ways because it is easier to than not, but, given that they do, the listener is provided with useful information. One possible other example is declination--the tendency for the fundamental frequency of the voice to drift downward over the course of a coherent syntactic unit (e.g., Cooper & Sorenson, 1981). Other things equal, F_0 will decline during an expiration as the lungs deflate. Declination due to this effect is observed even in word sequences produced with no communicative intent (Sternberg, Wright, Knoll, & Monsell, 1980). Talkers tend to take breaths at major syntactic (or metrical) boundaries (e.g., Grosjean & Collins, 1979) so, other things equal, F_0 will rise there too.

Therefore, declination and resetting will tend to be deployed systematically even though the talker is essentially just letting declination happen during expiration. Because F_0 resetting is systematic, however, the listener can use it as redundant information demarcating major syntactic boundaries.

Of course, the whole account of declination may be more complicated (see, for example, Cooper & Sorenson, 1981, who think that it is much, much more complicated). It has been highly controversial whether declination can be seen as an automatic consequence of lung deflation or, instead must be seen as an intentional imposition by the talker (compare Cohen, Collier, & t'Hart, 1982; Cooper & Sorenson, 1981; Gelfer, Harris, Collier, & Baer, 1983). We guess that the near universality of declination across languages (see the review by Cooper & Sorenson) is explained by the observation that it is easier for the talker to exhibit declination on expiration

than not. However, because declination is informative and because listeners use the information (see, e.g., Breckinridge, 1977), talkers may on occasion use declination and resetting intentionally to provide information at a boundary whether the talker does not need to take a breath.

More generally, we hypothesize that the confluence of articulatory ease and perceptual redundancy may promote perpetuation of systematic deployment of various kinds of articulatory information in speech.

ACKNOWLEDGMENT

We thank Kristin Snow for her help in collecting and analyzing the data from several of the experiments. We also thank Carole Beal and George Wolford for their comments on an earlier version of the manuscript and George Wolford for help with the statistical analysis of data from Experiment 1. The research was supported by NICHD Grant HD-01994 to Haskins Laboratories.

REFERENCES

Bolinger, D. (1963). Length, vowel, juncture. *Linguistics, 1,* 5-29.

Bolinger, D. (1981). *Two kinds of vowels, two kinds of rhythm.* Bloomington, IN: Indiana University Linguistics Club.

Breckinridge, J. (1977). *Declination as a phonological process.* Bell Laboratories Technical Memo, Murray Hill, NJ.

Chafe, W. (1974). Language and consciousness. *Language, 50,* 111-133.

Cohen, A., Collier, R., & t'Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch. *Phonetica, 39,* 254-273.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Cooper, W. E., & Sorenson, J. (1981). *Fundamental frequency in speech production.* New York: Springer-Verlag.

Fowler, C. A., Napps, S., & Feldman, L. (1985). Relations among regular and irregular morphologically-related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13,* 241-255.

Gelfer, C., Harris, K., Collier, R., & Baer, T. (1983). Speculations on the control of fundamental frequency declination. *Haskins Laboratories Status Report on Speech Research, SR-76,* 51-63.

Gough, P., Alford, A., & Holley-Wilcox, P. (1981). Words and context. In O. Tzeng & H. Singer (Eds.), *Perception of print* (pp. 85-102). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica, 36,* 98-114.

Housum, J. (1986). *Specification of given and new information in conversation.* Unpublished manuscript.

Hunnicutt, S. (1985). Intelligibility versus redundancy--conditions of dependency. *Language and Speech, 28,* 45-56.

Keillor, G. (1985). *Sylvester Krueger's desk.* Gospel Birds cassette tape, Minnesota Public Radio.

Kempley, S., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology, 73,* 441-454.

Koopmans-Van Beinum, F. J. (1980). *Vowel contrast reduction: An acoustic and perceptual study of Dutch vowels in various speech conditions.* Amsterdam: Academische Pers B. V.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 172-187.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668-682.

Pickett, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech, 6,* 151-164.

Sternberg, S., Wright, C., Knoll, R., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 507-534). Hillsdale, NJ: Lawrence Erlbaum Associates.

Zwicky, A. (1972). On casual speech. In P. Peranteau, J. Levi, & G. Phares (Eds.), *Papers from the eighth regional meeting of the Chicago Linguistics Society* (pp. 607-615). Chicago: Chicago Linguistics Society.

FOOTNOTES

*Journal of Memory and Language, 26, 489-504 (1987).

†Also Dartmouth College.

††Dartmouth College.

[1]Experiments 1-3 are replications of research by the second author performed as part of his Senior Honors project at Dartmouth College (Housum, 1986).

[2]We thank Kristin Snow for making these measurements.

[3]Housum (1986) did find a significant negative correlation between shortening and distance in his collection of spontaneous speech.

APPENDIX A

Excerpt from "Sylvester Krueger's Desk" by Garrison Keillor

Oh that smell about it, same blackboard, same portraits of Washington and Lincoln up front and center, up over the blackboard, Washington on the left, Lincoln on the right. Looking down on us all these years like an old married couple up there on the wall. I'd sit there at my desk, you know, bent over the paper trying to make big fat vowels so that the tops of them would just scrape the little dotted line. Make the tails of the consonants, the ps and the qs and the gs and fs so that they hung down. There I'd sit and memorize arithmetic tables and memorize state capitols and major exports of many lands. And whenever I was stumped, I'd always look up to see their pictures.

APPENDIX B

Methods for Topicality Rating Study

Subjects. Subjects were 10 students at Dartmouth College who participated for course credit. They were native speakers of English.

Materials. Subjects were given a typed transcription of the monologue, Sylvester Krueger's Desk. In addition, they received a rating sheet. On the sheet the 35 word pairs measured in Experiment 1 were listed. Next to each word were listed the pages and line numbers of its first two occurrences. In addition, there were slots for three topicality ratings.

Procedure. Subjects were run in groups of two to four. On arrival they were given copies of the seven-page transcription and they were asked to read it through quickly. As each subject finished, he or she was given a rating sheet and typed instructions. The instructions asked subjects to locate each relevant occurrence of a word in the passage and to rate the word's importance to the topic of its sentence. Next they were to rate the importance of the word's meaning to the topic of the monologue as a whole. Examples were provided from a different text to illustrate important and less important words in their respective sentences and passages.

Subjects reported no difficulty in following instructions. The session lasted about one half hour.

Word-initial Consonant Length in Pattani Malay*

Arthur S. Abramson†

Pattani Malay has distinctive length in all word-initial consonants. Earlier work showed that variations in closure-duration yield perceptual shifts between "short" and "long" phonemes for all sentence-medial intervocalic consonants but only for sentence-initial consonants with acoustic excitation before the release. For words, however, with initial voiceless closures but no pre-release excitation, which are identified well in isolation, where are the cues to the "length" distinction? In the belief that the underlying mechanism is the temporal control of closure, two hypotheses are tested here acoustically: (1) For all consonants, the closure-durations differentiate the short and long categories. (2) The ratio of the amplitude of the first syllable to the second syllable is greater in disyllabic words with long plosives than in those with short plosives.

BACKGROUND

The use of time and timing (Lehiste, 1970; Lisker, 1974) for phonological distinctions is still an important topic for research. This study tries to shed further light on the acoustic bases of length contrasts in which the relative durations of vocalic and consonantal gestures seem to have a distinctive function. Insofar as it might be a phonetic matter rather than an abstract phonological one, the question of whether to treat long segments as "geminates" will not be handled here.

Treatments of phonemic consonant length usually discuss intervocalic consonants, as in Estonian and Italian, where it is easy to show the physical reliability and perceptual relevance of durational differences in closures and constrictions. A language with this distinction in word-initial, and thus potentially, utterance-initial position, is rare.

THE LANGUAGE

Pattani Malay, spoken by some 600,000 ethnic Malays in southeastern Thailand, has a length-distinction for all consonants in word-initial position (Chaiyanara, 1983). (The language was first called to my attention by Christopher Court and Jimmy G. Harris.) Here are some word-pairs with the contrast:

/makɛ/	'to eat'	/m:akɛ/	'to be eaten'
/lamaʔ/	'late'	/l:amʔ/	'to make late'
/siku/	'elbow'	/s:iku/	'hand-tool'
/dʒalɛ/	'way'	/dʒ:alɛ/	'to walk'
/butɔ/	'blind'	/b:utɔ/	'a kind of tree'

All of the foregoing examples have acoustic excitation during their closures or constrictions, but there is none in the voiceless unaspirated plosives, as in these examples:

/tɕuɣi/	'to rob'	/tɕ:uɣi/	'robber'
/tawa/	'bland'	/t:awa/	'to show wares'

Recent work (Abramson, in press) has shown the power of closure-duration as an acoustic cue to the short-long distinction. Incremental shortening of acoustically excited closures yields perceptual shifts from long to short consonants. Voiceless plosives with their silent closures can be tested only in utterance-medial intervocalic slots; there, shortening or lengthening a silent gap induces shifts.

GOALS

The justification for the perceptual experiments (Abramson, in press) was impressionistic observations of length and a small body of instrumental measurements. The first goal here was to determine the statistical reliability of closure-duration as a differentiator of the categories. The second goal was to explore the possible role of overall amplitude in the distinction. That is, for utterance-initial voiceless plosives, something other than audible differences in closure durations must convey the distinction. Although other acoustic features, such as fundamental-frequency shifts and formant-transition rates, are not ruled out, the hypothesis considered here was that the aerodynamic consequences of the apparent articulatory mechanism would cause a higher amplitude upon the release of a long plosive.

DATA

Recordings were made of several native speakers, but only those of one man, PMC, were analyzed for this report. Minimal pairs of disyllabic words, two tokens of each, were elicited in isolation and in a carrier sentence. These utterances were digitized for measurement in a waveform editing program and for spectral analysis.

DURATION

The durations of all closures and constrictions were measured for all utterance-initial consonants—except, of course, for the voiceless ones—and all utterance-medial consonants. This was done by examining the waveforms for acoustic signs of forming and releasing obstructions in the supraglottal vocal tract; these were mainly release bursts and sudden changes in amplitude. Occasional difficult cases were checked against spectrograms. The data are summarized in Figure 1.

An analysis of variance showed duration to be highly significant for initial consonants, $F(1, 26) = 49.40$, $p < 0.0001$, and medial consonants, $F(1, 42) = 185.19$, $p < 0.0001$. To measure durations of initial voiceless closures would require either a direct look at articulation or, perhaps, measurements of buccal air pressure. The robustness of the difference for medial voiceless plosives, in conformity with the graphs for the medials in Figure 1, and the data in both positions for all other consonants, suggest the high probability of a closure-duration difference for initial voiceless plosives too.

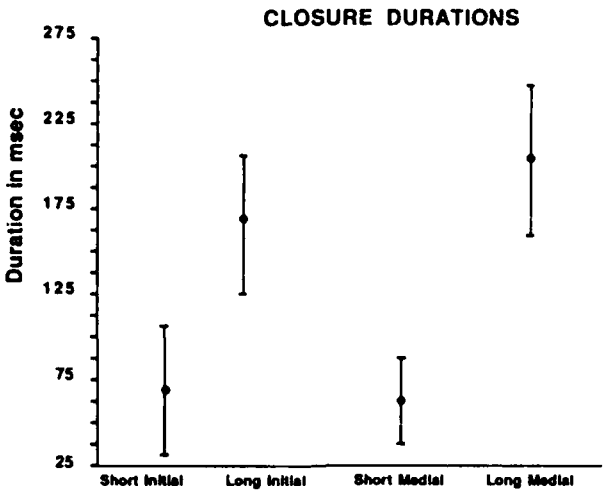


Figure 1. Means and one-standard-deviation error bars for Speaker PMC. Initial: C, n = 28; C:, n = 28. Medial: C, n = 44; C:, n = 44.

AMPLITUDE

Since the major concern was with initial voiceless plosives, measurements of amplitude were limited to isolated words. Pilot work with rise time, peak value, and average amplitude of the first syllable relative to the second gave useful results only with the third method.

A program with variable window-settings, designed by Richard S. McGowan, was used to derive the average root-mean-square (RMS) amplitude of each syllable in the disyllabic words recorded. (Apparently, monosyllabic words are rare.) The results are given in Table 1.

TABLE 1

Means and Standard Deviations for RMS Amplitudes in dB

		Short Consonants			*Long Consonants*		
Type	Syl. #	n	M	SD	n	M	SD
PLOSIVES							
Voiceless	1	16	47.5	3.0	16	51.0	2.2
	2	16	45.0	2.8	16	45.0	2.3
Voiced	1	14	46.8	3.9	14	49.5	3.5
	2	14	43.3	3.4	14	44.4	2.9
CONTINUANTS	1	14	45.1	3.9	14	48.1	2.8
	2	14	46.1	3.4	14	46.9	3.4

As expected, the most promising set of data in Table 1 is for the voiceless plosives (stops and affricates). In the analysis of variance of the underlying data, the interaction between consonant length and syllable approached significance: $F(1, 14) = 4.36$, $p = 0.056$. Indeed, post-hoc simple-effects tests showed that the difference between the short and long consonants with respect to amplitude-ratio is strongly significant: $F(1, 14) = 11.037$, $p = 0.005$. Although the continuants (nasals, laterals, and fricatives) showed a slight tendency in the same direction, the effect was not statistically significant. Compared with the continuants, the voiced plosives present a stronger case in the simple-effects test: $F(1, 24) = 4.24$, $p = 0.05$. With its greater number of degrees of freedom, however, this category underwent a more powerful test than the voiceless plosives and yielded a weaker although significant effect.

CONCLUSION

That the phonemic distinction between "short" and "long" Pattani Malay consonants is based on the quantitative feature of articulatory timing is abundantly clear from the data of Figure 1. Indeed, the perceptual efficacy of closure-durations has been demonstrated for medial position and for initial consonants with audible excitation (Abramson, in press). (Of course, the value of this cue has been demonstrated for at least medial position in some other languages [e.g., Lahiri & Hankamer, 1986].)

Even if, as seems likely, the underlying mechanism for this length distinction is articulatory timing, there may nevertheless be more than one acoustic cue involved. That is, temporal control of closures and constriction, intersecting with states of the glottis, may engender not only varying spans of silence or appropriate sound but also, perhaps, variations in air flow and pressure with certain acoustic consequences. The data in Table 1 show that for long voiceless initial plosives the average RMS amplitude is significantly higher in the first syllable than the second. There is also a significant but somewhat smaller effect for voiced plosives. We may speculate that although both categories involve complete momentary obstruction of the oral air flow, the presumed greater impedance at the larynx for the voiced plosives lessens the effect. For the continuants, however, which always have a by-pass for the air, there is no effect.

The amplitudes of PMC's embedded words remain to be measured. In the meantime, a cursory look at the productions of three other native speakers of the language seems to support the findings. Their utterances, too, will have to be measured. Finally, to round out the first experiments on perception (Abramson, 1986), the plan is to produce stimuli with controlled variations in amplitude on disyllables.

ACKNOWLEDGMENT

The work was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Mr. Paitoon Masmintra Chaiyanara of The Prince of Songkhla University, Pattani and Dr. Theraphan L. Thongkum of Chulalongkorn University, Bangkok for their help and to their institutions for their warm hospitality. The advice of Dr. Richard S. McGowan and Professor Leonard Katz was most helpful.

REFERENCES

Abramson, A. S. (in press). Distinctive length in initial consonants: Pattani Malay. *Journal of the International Phonetic Association.* [Also in this *Status Report*, SR-91.]

Chaiyanara, P. M. (1983). *Dialek Melayu Patani dan Bahasa Malaysia: Satu Kajian Perbandingan dari segi Fonologi, Morfologi dan Syntaksis.* Master's thesis, University of Malaya.

Lahiri, A., & Hankamer, J. (1986). Acoustic properties of geminate consonants. *Journal of the Acoustical Society of America, 80*, S62 (A).

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press

Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok et al. (Eds.), *Current trends in linguistics* (Vol. 12, pp. 2387-2418). The Hague: Mouton.

FOOTNOTES

*Paper presented at the 11th International Congress of Phonetic Sciences, held 1-7 August 1987 in Tallinn, Estonia, U.S.S.R., and published in the *Proceedings XIth ICPhS, Vol. 6*, pp. 68-70, Tallinn 1987.

†Also University of Connecticut

The Perception of Word-initial Consonant Length: Pattani Malay*

Arthur S. Abramson†

The most salient physical manifestation of phonemically distinctive consonant length is the duration of the closure or constriction of the short consonant relative to that of its long counterpart. The contrast is rare in languages of the world in word-initial, and thus potentially utterance-initial, position. Perception in this position would seem to depend upon the audibility of closure excitation. This is plausible for nasals, laterals, and fricatives. The closures of voiced stops, however, may have only low-amplitude excitation, while voiceless stops have none. Pattani Malay was investigated to find out how robust the length feature is in perception. Listening tests yielded good differentiation of the two length classes for isolated words, with a lesser effect for voiceless stops. Experiments with incrementally lengthened short closures and shortened long closures confirmed the sufficiency of duration as a cue. For the voiceless stops, these experiments could be run only in intervocalic position.

In experimental phonetic research, we often come across cases of multiple perceptual cues to a phonemic distinction, even though this distinction may be seen traditionally as dependent on some single phonetic feature. The question arises as to the relative power of these cues: How equally do they share the burden of communicative relevance? For example, among the several cues that emanate from the timing of the valvular action of the larynx (Abramson, 1977; Lisker & Abramson, 1965), fundamental-frequency perturbations (House & Fairbanks, 1953) have been shown by some studies, apparently starting with Haggard, Ambler, and Callow (1970) and Fujimura (1971), to help in the perceptual differentiation of voiced and voiceless stop consonants; however, recent work (Abramson & Lisker, 1985) suggests that this cue has very limited efficacy compared with other acoustic consequences of voice timing.

The present study is meant to combine the foregoing interest with an attempt to shed further light on the perceptual basis of length contrasts as found in many languages in which phonologically distinctive functions are borne by the relative durations of vowels (Abramson, 1962; Lehiste, 1970) or of the closures and constrictions of consonants (Lehiste, 1970). Discussions of the feature of consonant length usually focus on intervocalic consonants, as in Italian and Estonian. In such cases, it is easy to demonstrate the existence of differences in consonant-closure duration and their perceptual relevance. What is uncommon is to find a language with such a distinction in word-initial, and thus potentially utterance-initial, position. In length distinctions in any context, it is not unlikely that other acoustic features will covary with the duration of the relevant span of speech. Perhaps some of these concomitant features serve as cues together with duration or, in certain circumstances, instead of duration.

Pattani Malay, the dialect of Malay spoken by some 600,000 ethnic Malays in southeastern Thailand, has a length distinction for all consonants in word-initial position (Chaiyanara, 1983), thus for all consonant classes of the language. The language was first called to my attention by the fieldwork of Christopher Court and Jimmy G. Harris. Here are some word pairs with the contrast:

/labɔ/	'to make a profit'	/l:abɔ/	'spider'
/makɛ/	'to eat'	/m:akɛ/	'to be eaten'
/siku/	'elbow'	/s:iku/	'hand-tool'
/bulɛ/	'moon'	/b:ulɛ/	'many months'
/kato?/	'to strike'	/k:ato?/	'frog'

For all consonants with acoustic excitation of any kind during the closure or constriction, it is obvious that closure duration alone could be enough to differentiate the members of each pair in both production and perception. The first four pairs of examples are of that type, while the fifth pair, with voiceless unaspirated stops, is not. That is, in pairs of the last type, the difference in the closure durations appears only as shorter or longer medial silent gaps when the words are embedded in utterances.

For the two speakers examined in preparation for this perceptual study, the closure durations of the long consonants in both initial and medial position are on average three times longer than those of the short consonants. A more detailed presentation of duration ratios is given elsewhere (Abramson, 1987). For the voiceless stops and affricates, measurements of duration can be made, of course, only in utterance-medial position.

If indeed the timing of the closure is the articulatory mechanism underlying the distinction, wherever the difference in closure duration is audible, it ought to be a sufficient auditory cue. In intervocalic contexts, the abrupt spectral shifts between the vocalic portions of the signal and the consonantal closure rather clearly define the acoustic span correponding with the interval of the closure. It is in utterance-initial position that the question arises as to the sufficiency of this cue. One might expect that nasals, laterals, and fricatives with high radiation of sound during closure would be as well distinguished by listeners in initial position as in intervocalic position. Voiced stops and affricates might be an intermediate case. There is audible glottal pulsing during the closures, but its amplitude may be too low for completely reliable differentiation based on two values of voicing lead. Voiceless plosives, however, with no acoustic excitation in the closure, must surely be differentiated in initial position by other cues, possibly ones that are themselves a function of the temporal articulatory feature. These might be intensity of stop-release burst, rate of formant transitions, or fundamental-frequency perturbations. A separate study, started somewhat later, has shown that in disyllabic words, if the first syllable begins with a long stop consonant, the amplitude of that syllable is significantly greater than that of the second syllable, but that is not so if the word begins with a short consonant (Abramson, 1987).

EXPERIMENTS

EXPERIMENT 1

To establish a baseline against which to do manipulative experiments, it was first necessary to assess the perceptual robustness of the length distinction. Thirty-two

Pattani Malay words forming 16 short-long pairs were recorded in isolation in two random orders by two women who, although also fluent in Thai, were native speakers of Pattani Malay and lifelong speakers of it. The resulting two test orders, each one containing two tokens of each word, were played through headphones to 21 students, all native speakers, on the Pattani campus of the Prince of Songkhla University. Because this language has not been reduced to writing for popular use, it was necessary to provide the test subjects with answer sheets with a pair of possible responses written in Thai next to each item number. The possible responses were short Thai glosses for the Malay words. For each item, the subject encircled whichever of the two glosses he or she thought appropriate for the spoken word. A member of the faculty,[1] also a native speaker of the language, gave the instructions in Pattani Malay on procedure and went over all the glosses to be sure that there were no misunderstandings. This method was used for all subsequent tests and seemed to cause no trouble.

The results of the baseline experiment are given in Figure 1. As can be seen, the nasals, laterals, and fricatives were all identified quite well. Next, but not much lower in percentages correct, come the stops. Worst are the affricates, especially the voiceless ones, which were not labeled much better than chance. This led me to decide not to work further with the affricates in the present investigation.

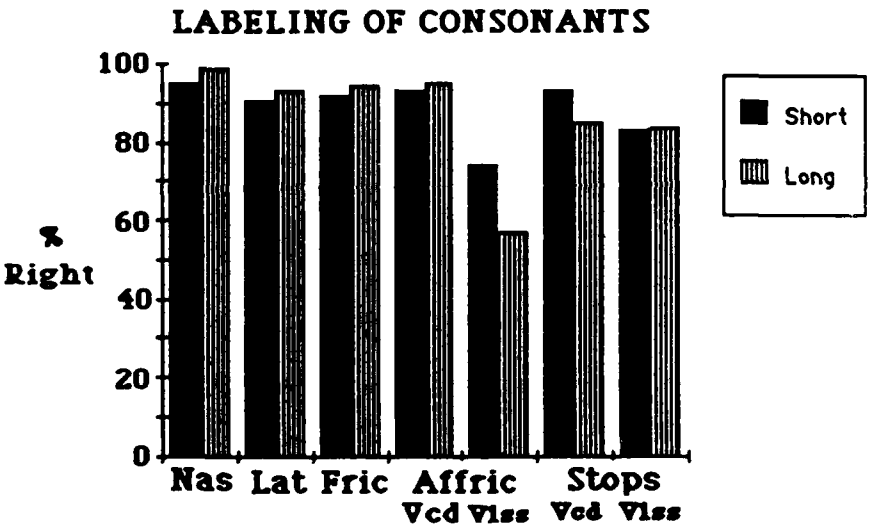


Figure 1. Experiment 1. Identification of natural utterances of isolated words beginning with short and long consonants.

EXPERIMENT 2

For this stage of my perceptual research on the distinction, I have limited my manipulations to the obvious variable of closure duration.[2] If it is indeed a major cue to the distinction, we ought to be able to bring about a shift in percept by big enough changes in closure duration. Wishing to start with consonants with very audible voicing during closure, I chose a pair of words distinguished by short and long /l/: /labɔ/ 'to make a profit' versus /l:abɔ/ 'spider.' By means of wave-form editing, I shortened the initial lateral resonance of the isolated long member of the pair from its original duration of 183 ms to 63 ms in 12 10-ms steps. Thus, the shortest variant was shorter than the original short /l/ of 72 ms, which itself was not

used in the experiment. I made two more stimuli by cross-splicing the lateral resonances between the two original words. That is, I removed the lateral resonances, replacing the original long one with the short one and the original short one with the long one. The resulting 15 stimuli were recorded three times each into a random order for presentation to the 21 subjects.

The results for the shortened /l:/ series, without the cross-spliced stimuli, are given in Figure 2. The range of closure durations is shown along the abscissa, and the labeling percentages along the ordinate. The crossover zone between the two categories is centered around 100 ms. Clearly, duration is a sufficient and powerful cue to the distinction. The cross-spliced stimuli were heard virtually 100% of the time as the original words from which the resonances had been cut; this suggests that features of the constriction release had little or no cue value.

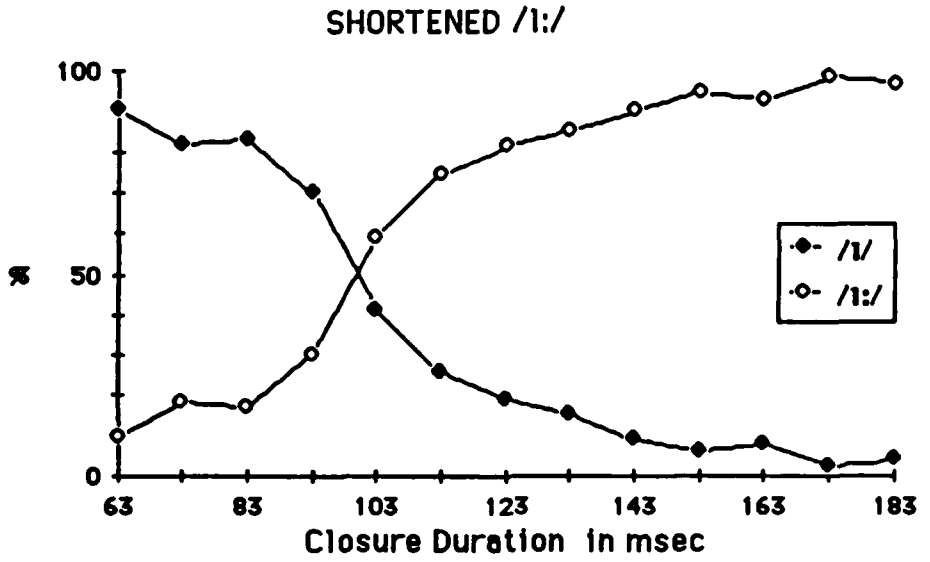


Figure 2. Experiment 2. Identification functions for original long /l:/ and shortened variants. The responses to the cross-spliced stimuli are not included.

The next two experiments examined the crucial case of voiceless stops, which, of course, could be subjected to closure-duration changes by the method of wave-form editing only in intervocalic position in a neutral Pattani Malay carrier sentence suitable for the two words in question. The words chosen for both experiments were /paka/ 'to use' and /p:aka/ 'usable.' The plan was to shorten the long consonant and lengthen the short consonant to test, once again, the effects of variation in closure duration and, indirectly, effects of stop release.

EXPERIMENT 3

In this experiment, I shortened the closure of long /p:/ from its original duration of 182 ms in 14 10-ms steps in a carrier sentence. The shortest variant at 42 ms was a bit shorter than the original short /p/ of 47 ms, which itself was not used in the experiment. The resulting 15 stimuli were recorded three times each into a random order for presentation to the 21 subjects in the carrier sentence.

The results of Experiment 3 are given in Figure 3. The sufficiency of relative duration as a cue to the length distinction is demonstrated here for voiceless intervocalic stops. The unexplained leveling of the responses for the items at 92 and

102 ms may well be due to an artifact that remains to be uncovered. If we ignore the latter, the 50% crossover point is at 104 ms; a curve-fitting procedure, or better, a replication with new stimuli might yield a slightly earlier crossover.

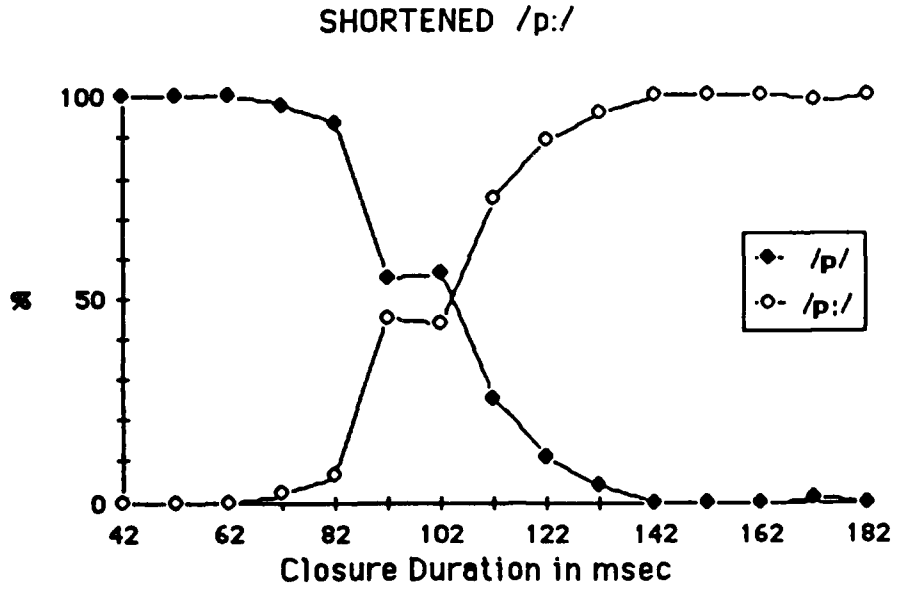


Figure 3. Experiment 3. Identification functions for original long /p:/ and shortened variants.

EXPERIMENT 4

This experiment differed from Experiment 3 in that /paka/, the short member of the pair, was the starting point. I lengthened the closure of short /p/ in the carrier sentence from its original duration of 47 ms in 14 10-ms steps. The longest variant was 187 ms, a bit longer than the original long /p:/ of 182 ms, which itself did not appear in the experiment. This simply required using our waveform-editing program to add increments of time to the middle of the closure gap. Again, the resulting 15 stimuli were recorded three times each into a random order and played to the 21 subjects for identification as words in the carrier sentence.

The results of Experiment 4 are given in Figure 4. Not surprisingly, the importance of closure duration is again apparent for intervocalic voiceless stops. Here, however, the perceptual cross-over point is at 120 ms, considerably later than the one in Experiment 3, with or without the possible artifact in the latter. An analysis of variance showed the large difference between the two crossover points to be highly significant, $F(1, 20) = 27.48$, $p < 0.0001$. This implies the probable efficacy of one or more cues concomitant with that of closure duration.

EXPERIMENT 5

The goal of this last experiment was to explore the intermediate condition, that of the presence of quasiperiodic acoustic excitation during the closure but at a low amplitude. The word pair chosen was /gamɔʔ/ 'approximately' vs. /g:amɔʔ/ 'to be shy.' I shortened the closure of long /g:/ in a citation form of the word from its original duration of 197 ms in 15 10-ms steps. The shortest variant at 47 ms was just under the 50-ms duration of the closure of the original short /g/, which itself was not used in the experiment. The 16 resulting stimuli were recorded three times each into a random order and played to the 21 subjects for identification.

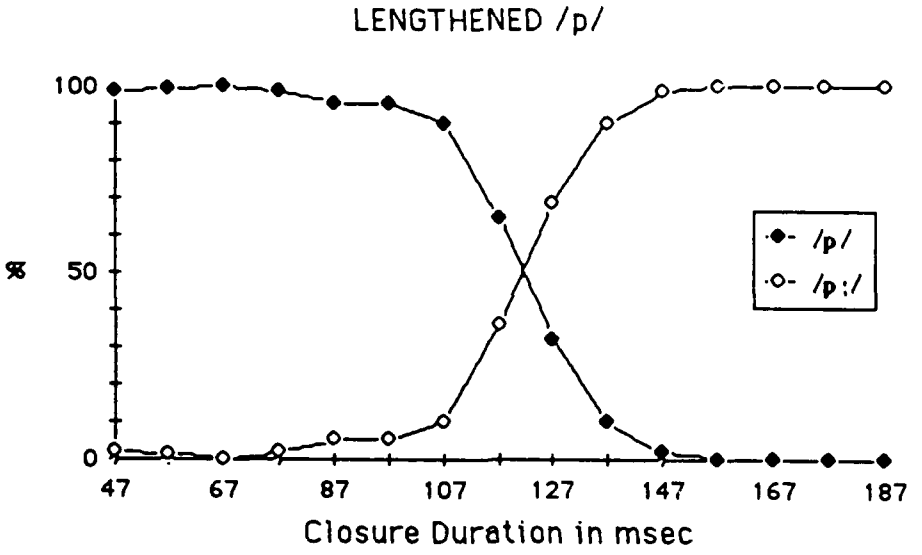


Figure 4. Experiment 4. Identification functions for original short /p/ and lengthened variants.

The results of Experiment 5 are given in Figure 5. The middle of the crossover zone between the two percepts is at 97 ms. Clearly, even with just low-amplitude voicing in the closures of initial stops, relative duration is a sufficient cue.

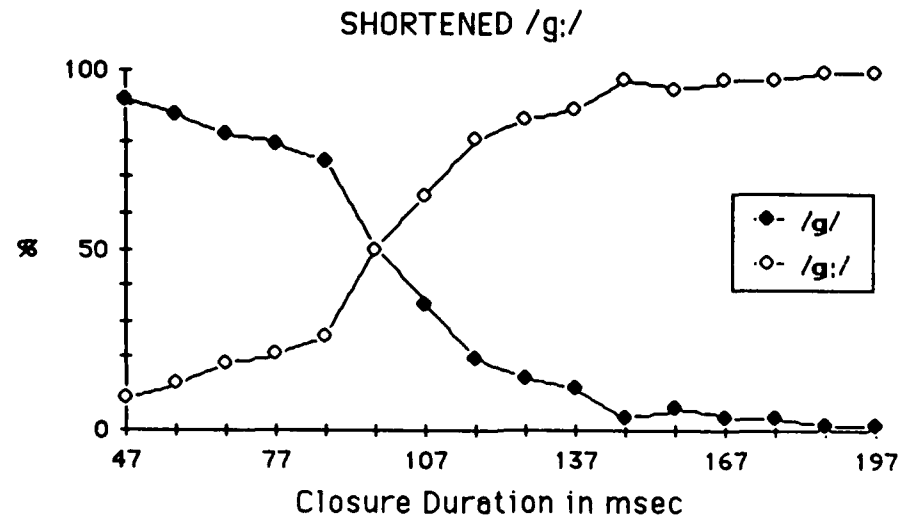


Figure 5. Experiment 5. Identification functions for original long /g:/ and shortened variants.

CONCLUSION

The word-initial consonants of Pattani Malay, except perhaps for the voiceless affricates, can be identified well as to length category in utterance-initial position. Perceptual experiments reveal the power of closure duration as a sufficient cue in medial and, when audible, initial position for the distinction between "short" and "long" consonants.

The early crossover at 33% of the durational difference between utterance-initial /g/ and /g:/ seems comparable to the crossover at 31% of the difference between /l/ and /l:/, as contrasted with 44% in Experiment 3 and 52% in Experiment 4, both of the latter for medial voiceless stops. With so few experiments so far, we can only speculate that for utterance-initial voice-excited closures or constrictions, there is some psychoacoustic threshold below which a long consonant cannot be heard.

The results of Experiments 2 and 3 imply that for the length distinction in voiceless medial stops there is another cue, if only a secondary one, in addition to relative duration. Obviously, given the results of Experiment 1, some such feature must be at work as the primary cue in utterance-initial position. Acoustic analysis (Abramson, 1987) has shown relative amplitude to be the most promising candidate. Perceptual testing of this hypothesis is the next thing to be done.

ACKNOWLEDGMENT

The work was supported by NICHD Grant HD–01994 to Haskins Laboratories. The excellent help of two colleagues in Thailand, Mr. Paitoon Masmintra Chaiyanara of the Prince of Songkhla University, Pattani and Dr. Theraphan L. Thongkum of Chulalongkorn University, Bangkok, enabled me to do the fieldwork for this study. I also wish to thank both institutions for their warm hospitality.

REFERENCES

Abramson, A. S. (1962). *The vowels and tones of Standard Thai: Acoustical measurements and experiments*. Bloomington: Indiana University Research Center in Anthropology, Folklore, and Linguistics.

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34*, 295-303.

Abramson, A. S. (1986). Distinctive length in initial consonants: Pattani Malay. *Journal of the Acoustical Society of America, 79*, S27 (Abstract).

Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. *Proceedings of the XIth International Congress of Phonetic Sciences* (pp. 68-70). Tallinn: Academy of Sciences of the Estonian S.S.R.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F_0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). New York: Academic Press.

Chaiyanara, P. M. (1983). *Dialek Melayu Patani dan Bahasa Malaysia: Satu Kajian Perbandingan dari segi Fonologi, Morfologi dan Syntaksis*. Master's thesis, University of Malaya.

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jørgensen* (pp. 221-231). Copenhagen: Akademisk Forlag.

Haggard, M. P., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America, 47*, 613-617.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America, 25*, 105-113.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lisker, L., & Abramson, A. S. (1965). Stop categorization and voice onset time. *Proceedings of the Vth International Congress of Phonetic Sciences* (pp. 389-391). Basel: S. Karger.

FOOTNOTES

*To appear in the *Journal of the International Phonetic Association*. This is a revised and expanded version of a paper read at the 111th Meeting of the Acoustical Society of America, Cleveland, Ohio, May 12-16, 1986 (Abramson, 1986).

†Also the University of Connecticut.

[1]Mr. Paitoon Masmintra Chaiyanara, who is thanked in the Acknowledgment, also went over all the words beforehand for their authenticity and for the accuracy of the glosses. He is engaged with colleagues in compiling a trilingual (Pattani Malay-Standard Malay-Thai) dictionary of the language.

[2]The results of recent analytic work (Abramson, 1987) call for perceptual experiments with relative amplitude as the variable.

Perception of the [m]-[n] Distinction inVC Syllables*

Bruno H. Repp and Katyanee Svastikula†

This study complements earlier experiments on the perception of the [m]-[n] distinction in CV syllables (Repp, 1986, 1987). Six talkers produced VC syllables consisting of [m] or [n] preceded by [i,a,u]. In listening experiments, these syllables were truncated from the beginning and/or from the end, or waveform portions surrounding the point of closure were replaced with noise, so as to map out the distribution of the place of articulation information for consonant perception. These manipulations revealed that the vocalic formant transitions alone conveyed about as much place of articulation information as did the nasal murmur alone, and both signal portions were about as informative in VC as in CV syllables. Nevertheless, full VC syllables were less accurately identified than full CV syllables, especially in female speech. The reason for this was hypotesized to be the relative absence of a salient spectral change between the vowel and the murmur in VC syllables. This hypothesis was supported by the relative ineffectiveness of two additional manipulations meant to disrupt the perception of relational spectral information (channel separation or temporal separation of vowel and murmur) and by subjects' poor identification scores for brief excerpts including the point of maximal spectral change. Whereas in CV syllables the abrupt spectral change from the murmur to the vowel provides important additional place of articulation information, for VC syllables it seems as if the formant transitions in the vowel and the murmur spectrum functioned as independent cues.

INTRODUCTION

In recent studies, Kurowski and Blumstein (1984) and Repp (1986, 1987) have investigated the perception of the [m]-[n] distinction in natural CV syllables. Their results have shown that, for prevocalic nasal consonants, place of articulation information is generally contained both in the spectrum of the nasal murmur and in the vocalic formant transitions following the point of release. In addition to combining these two separate sources of information, however, listeners derive information from the spectral relationship between the two signal portions, which appears to be a crucial cue for the [m]-[n] distinction in the context of front vowels such as [i]. Kurowski and Blumstein (1984, 1987) hypothesized the existence of a single auditory property for place of articulation representing the spectral change from the murmur into the vowel, and both they and Repp (1986) speculated about the possible role of auditory short-term adaptation caused by the murmur in establishing or enhancing this distinctive auditory property at vowel onset. The most recent perceptual results (Repp, 1987), however, suggest that the perception of this spectral relationship does not depend on peripheral auditory enhancement, at least not under favorable listening conditions.

The present study complements the CV syllable experiments of Repp (1986, 1987) by examining the perception of the [m]-[n] distinction in VC syllables using similar

methods. (To facilitate comparisons with the earlier data, the nasal consonant [ŋ], which occurs only in postvocalic position in English, was not included.) As in prevocalic nasals, the place of articulation of postvocalic nasals is conveyed by the vocalic formant transitions and the nasal murmur, occurring in reverse order. There are several important differences, however, which make a comparison interesting.

First, final nasal consonants may be released, and if so, the release transient (which may continue into a brkef neutral vowel) contains salient additional place of articulation cues. Clearly, to compare the perception of initial and final nasals, only final nasals without releases should be considered. Nevertheless, the omission of an additional (however optional) piece of information may entail some loss in intelligibility.

Second, there are two reasons for expecting the spectral relationship between vowel and murmur to be less salient perceptually in VC than in CV syllables. One reason is that, because the murmur *follows* the vowel containing the formant transitions ("vowel" is used here to denote the signal portion preceding the point of closure in a VC syllable), it cannot have any auditory adaptation effect on the vowel. Although adaptation caused by the vowel may modify the auditory representation of the murmur, it seems unlikely that this peripheral interaction would enhance place of articulation information, since it would only attenuate the already weak higher formants of the murmur, which are continuous with the formants at vowel offset. Thus, one process hypothesized to establish relational spectral information for place of articulation (Kurowski & Blumstein, 1984) presumably does not operate here. A second reason is that the transition between vowel and murmur in VC syllables is not as abrupt as the murmur-vowel transition in CV syllables (cf. Kurowski & Blumstein, 1987).

Vowels preceding nasal consonants are commonly nasalized, more so than following vowels (see, e.g., Ali, Gallagher, Goldstein, & Daniloff, 1971; Ostreicher & Sharf, 1976), and this anticipatory opening of the velar port reduces the spectral contrast between the vowel and the murmur. Also, Schouten and Pols (1979) have suggested that, while prevocalic nasal murmurs contain no formant transitions, formant movements may extend from a vowel into a following murmur, suggesting that articulatory adjustments continue after oral closure. This would also contribute to making the spectral change less abrupt.

Third, the perceptual contributions of the vowel and murmur components themselves may also differ between CV and VC syllables. The vocalic formant transitions of final nasals are not mirror-images of those of initial nasals, and in fact may be more distinctive (Broad & Fertig, 1970). Perceptual experiments with truncated CV and VC syllables containing stop consonants have suggested that VC transitions provide stronger place of articulation cues than CV transitions (Ohde & Sharf, 1977, 1981; Pols & Schouten, 1978, 1981). It remains to be seen whether this is also true for nasal consonants. Final nasals also have longer murmurs than initial ones (Malécot, 1956). Although murmur duration as such does not seem to have much of an effect on intelligibility (Repp, 1987), the possible presence of formant transitions in the murmur and its terminal position in the utterance may give it greater salience in VC than in CV syllables.

These considerations lead to the prediction that, although identification accuracy may be lower for (unreleased) final than for initial nasals in full syllables, the vowel and murmur components by themselves should be at least as informative in VC as in CV syllables. This paradoxical situation could arise because the spectral change between vowel and murmur is less important perceptually in VC syllables, so that the place of articulation information derives from two independent cues, as it were, without any additional "relational term" in the perceptual equation.

The only previous study in the readily accessible literature that compared the perception of nasal consonants in natural CV and VC syllables was conducted by Malécot (1956). It included [m, n, ŋ] in the context of a single vowel, [æ], apparently produced by a single talker. Stimuli were constructed by cross-splicing murmurs and vowels; murmurs, but not vowels, were also resented in isolation. The results suggested that murmur cues were more salient in final than in initial position, but they did not permit any conclusions about the relative contribution of the vocalic formant transitions.

EXPERIMENT 1

Repp's (1986) waveform-editing study with CV syllables included five conditions, four of which were replicated here with VC syllables: progressive truncation from the beginning, progressive truncation from the end, presentation of brief excerpts from the vicinity of the point of closure (corresponding to the point of release in CV syllables), and replacement of the same brief segments in the intact syllables with signal-correlated (i.e., envelope-matched) noise. The first two conditions served to determine the relative informativeness of the vowel and murmur portions in isolation, and the extent to which place of articulation information in each is located near the closure point. The other two conditions assessed the perceptual importance of the relationship between vowel and murmur spectra, and the extent to which that relational information rests on the availability of the point of maximal spectral change (the closure point). If that point is perceptually important, brief excerpts straddling the closure point should yield higher identification performance than excerpts from either side of that point, and replacement of these segments with noise should lead to lower identification scores than replacement of segments from within the vowel or the murmur.

Methods

Talkers and Recording Procedure

Six native speakers of American English served as talkers, three males (AA, GK, and JS) and three females (CG, SN, and BT). Five were researchers or graduate students under 40 years of age; one (AA) was an experienced phonetician in his early sixties. Three of the talkers (AA, CG, BT) had also served in the earlier study on CV syllables (Repp, 1986); the other three talkers of that study were no longer available and had to be replaced.

The talkers were asked to produce the syllables [ɑm, im, um, ɑn, in, un] as naturally as possible. The recording was done in a sound-insulated booth using high-quality equipment.

Stimuli and Test Sequences

The basic set of stimuli included 36 syllables (6 talkers x 6 syllables). These syllables were low-pass filtered at 4.9 kHz, digitized at a 10 kHz sampling rate, and stored in separate computer files. The waveforms were then inspected to determine whether any syllables had a final release. Of the 36 tokens, 14 were found to be released (all utterances of female talkers CG and BT, and one each of JS and SN). In each of these tokens, the portion of the waveform including and following the final release was removed, so as to ensure homogeneity of the stimulus set and to facilitate comparisons with syllable-initial consonants. It was assumed that these tokens would be equivalent to originally unreleased ones; however, see Experiment 2 below for a detailed investigation of this issue.

Subsequently, a waveform editor was used to place seven markers ("cutpoints") in each file, as illustrated in Figure 1. The marker labeled "0" was placed at the onset of what was taken to be the first glottal cycle of the nasal murmur. This point (the closure point) was defined as a visible amplitude drop and/or a decrease in high-frequency oscillations in the waveform. For reasons having to do with the ease of locating exact zero crossings (see Repp, 1986), the marker was placed at a downgoing zero crossing in male waveforms (Figure 1, upper panel), but at an upgoing zero crossing in female waveforms (lower panel). No perceptual consequences of this procedural difference were expected. The closure point could be determined with some confidence in [ɑ-] and [i-] syllables (see upper panel), but it was almost impossible to find in [u-] syllables (see lower panel). In these syllables, therefore, we made an "educated guess" based on the slope of the amplitude envelope, on the expected intrinsic duration of [u] as compared to [i] and [ɑ] (Peterson & Lehiste, 1960), and on listening carefully to the gated stimulus portions. The markers labeled -3, -2, -1, +1, +2, and +3 were placed at the onsets of glottal cycles preceding and following the "0" marker. The intermarker intervals will be referred to as "segments." There was one glottal cycle per segment for males (except for talker JS, whose relatively high fundamental frequency in three syllables suggested having two glottal cycles per segment in those tokens) and two for females. The average durations of the intermarker intervals, calculated over the -3 to +3 range, and the corresponding average fundamental frequencies in the vicinity of the closure point for the six talkers were as follows: 10.3 ms, 97 Hz (AA); 9 ms, 112 Hz (GK); 10 ms, 146 Hz (JS); 9.2 ms, 217 Hz (CG); 11 ms, 182 Hz (SN); 11.3 ms, 176 Hz (BT). A nominal segment duration of 10 ms will be assumed in discussing the results.

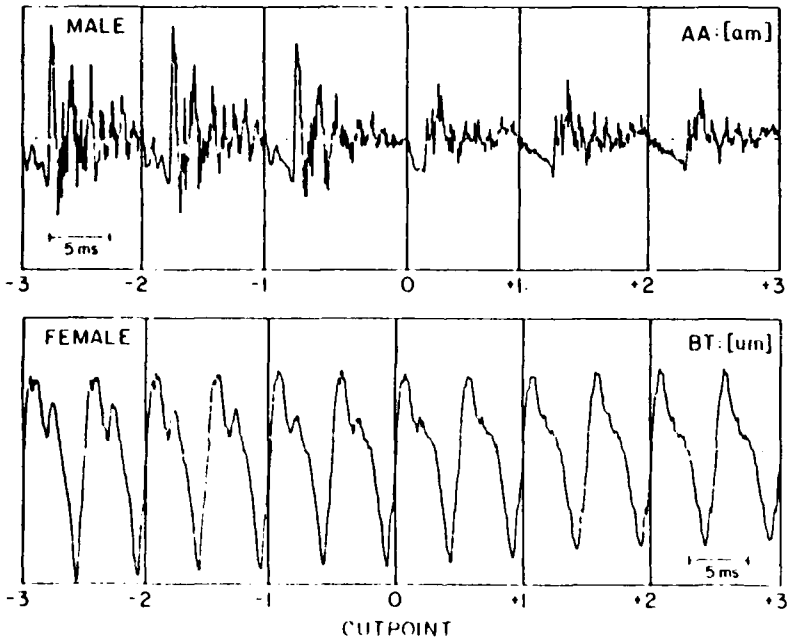


Figure 1. Oscillograms of the waveforms in the vicinity of the presumed point of closure for a male [ɑm] and a female [um], with cutpoint markers in place.

The stimuli, with cutpoint markers in place, were used to prepare four test tapes corresponding to four experimental conditions. Each test tape contained 7-8 test

sequences. Each test sequence consisted of the 36 individual syllables, modified as described below, in random order. The interstimulus interval was 3 seconds.

(a) Truncation from the beginning ("Murmurs"). There were eight test sequences in this condition. The first sequence contained the full syllables. The remaining seven sequences presented the stimuli starting at cutpoints -3, -2, -1, 0, +1, +2, and +3, respectively.

(b) Truncation from the end ("Vowels"). Eight test sequences were prepared for this condition. The first sequence contained the full syllables. The remaining sequences presented the stimuli ending at cutpoints +3, +2, +1, 0, -1, -2, and -3, respectively.

(c) Extraction of brief segments ("Excerpts"). This tape contained seven test sequences containing the following excerpts: -3/+3 (i.e., from cutpoint -3 to cutpoint +3), -2/+2, -1/+1, -2/0, 0/+2, -3/-1, and +1/+3. Therefore, the duration of the stimuli was about 60 ms in the first sequence, 40 ms in the second sequence, and 20 ms in the remaining sequences. The segments in sequences 1-3 straddled the closure point, whereas those in sequences 4 and 6 came from within the vowel and those in sequences 5 and 7 came from within the murmur.

(d) Replacement of segments with signal-correlated noise ("SCN"). There were seven test sequences in this condition, each containing full syllables in which the +1/+3, -3/-1, 0/+2, -2/0, -1/+1, -2/+2, and -3/+3 segments had been replaced with signal-correlated noise. (The order is reversed with respect to the Excerpt tape.) The duration of the noise in the syllables thus was about 20 ms in sequences 1-5, 40 ms in sequence 6, and 60 ms in sequence 7. The noise was generated by a computer program from specified segments within each waveform by randomly reversing the polarity of digital sampling points with a probability of .5 (Schroeder, 1968). By this method, the amplitude envelope of the original signal was maintained, but the spectral information was destroyed.

Subjects and Procedure

Twelve native speakers of American English served as listeners. They were paid student volunteers with reportedly normal hearing.

The test tapes were played to the subjects binaurally at a comfortable volume over TDH-39 earphones in a quiet room. There were one or two subjects per session, which lasted about 90 minutes. The Excerpts tape was always presented last, since it was considered to be the most difficult test due to the short duration of the stimuli. The other three conditions were presented in all six possible orders, with two subjects for each. The order of test sequences within each condition was fixed roughly according to progressive difficulty, as described above.

The subjects were asked to judge whether a stimulus was derived from a syllable ending with [m] or [n], by writing down /m/ or /n/ for each stimulus. If no nasal consonant was heard, a guess was to be made. Subjects were told that there were several talkers, and that there was an equal number of [-m] and [-n] syllables. One subject was replaced because his identification of the unaltered syllables was at chance level.

Data Analysis

The (untransformed) data were analyzed in two types of analysis of variance (ANOVA): across subjects (averaged over talkers) and across talkers (averaged over subjects). Therefore, two F values will be reported for each effect tested, and only effects for which both F values are significant will be reported. Differences among individual syllables were assessed by including consonant and vowel as factors in the ANOVAs. In the analysis across talkers, talker sex was an additional factor.

Results and Discussion

Murmurs

The results of the Murmurs condition are shown in Figure 2 as the open circles. It may be noted, first, that the full syllables were not perfectly identified: The average score was only 88% correct, which contrasts with the near-perfect identification of unaltered CV syllables (Repp, 1986). Closer inspection of the data revealed that the female tokens were much more poorly identified (77% correct) than the male tokens (98% correct). Four female tokens were especially conspicuous in that they tended to be identified at or below chance accuracy. Three of them (CG's [ɑn] and [in], and BT's [um]) had been originally released, and listening to the original stimuli confirmed that they had been correctly articulated by the talkers. Thus the removal of the final release may have impaired the intelligibility of these syllables (however, see Experiment 2 below). The fourth "bad" token (SN's [um]) may have been poorly articulated. Even with these tokens omitted, however, the score for female utterances was only 90% correct.

As the vowel was cut back, performance declined to somewhat below 70% correct. The leveling off of identification performance in the vicinity of the "0" cutpoint confirms that this marker had been placed with reasonable accuracy. Further cutback of the murmur itself did not lead to any decline in identification scores. This is not surprising, since the murmurs were rather long in duration (236 ms on the average, ranging from 119 to 416 ms), so that removal of the initial 30 ms hardly made any difference. If there were any formant movements during this portion (as observed by Schouten & Pols, 1979), they were not perceptually salient. Identification of the isolated VC murmurs was quite comparable in accuracy to that of isolated CV murmurs (Repp, 1986), even though the latter were of much shorter duration. Since murmur duration has relatively little influence on identifiability within limits (Repp, 1986, 1987), it may be concluded that VC and CV murmurs convey about the same amount of place of articulation information. Malécot's (1956) observation that final murmurs are perceptually more salient than initial ones may hold only when conflicting transitions and murmurs are spliced together.

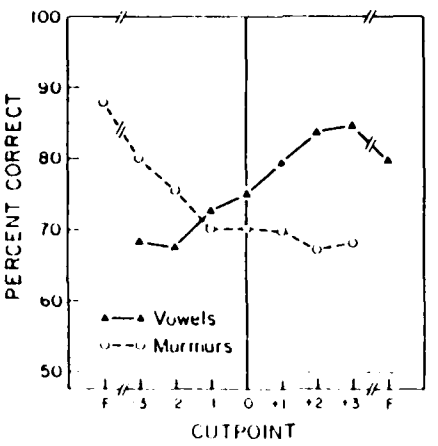


Figure 2. Percent correct identification for syllables in the Murmurs and Vowels conditions of Experiment 1 as a function of truncation. (F = full syllable.)

The scores for individual syllables, averaged over talkers, are shown in Figure 3. It is evident that [-m] syllables suffered much less from elimination of the vocalic formant transitions than did [-n] syllables. That is, isolated [m] murmurs were

identified more accurately than [n] murmurs. It is not clear whether this should be considered a response bias or a consequence of labial place of articulation information somehow being conveyed more strongly in murmurs (see also Malécot, 1956; Repp, 1987). It cannot have been entirely due to a response bias, however, because it depended on the original vocalic context: [(u)m] murmurs were less well identified than [(a)m] and [(i)m] murmurs (though there may have been one bad token of [um]), but [(a)n] murmurs were less well identified than [(u)n] and [(i)n] murmurs. Although most of these differences parallel those found with CV syllables (Repp, 1986), there is one striking difference: While [m(i)] murmurs were identified at chance level, [(i)m] murmurs were identified quite well (82% correct).

The ANOVAs showed the expected significant main effect of cutback: $F(7,77) = 14.02$, $p < .0001$; $F(7,28) = 12.17$, $p < .0001$. The consonant by cutback interaction was also significant, $F(7,77) = 4.63$, $p = .0002$; $F(7,28) = 3.88$, $p = .0045$, which confirms the trend of [-n] syllables to be harmed more by truncation than [-m] syllables. Furthermore, there was a significant vowel by consonant interaction, $F(2,22) = 30.40$, $p < .0001$; $F(2,8) = 5.96$, $p = .0260$, reflecting the fact that [m] syllables showed the opposite effects of vowel context ([a] > [i] > [u]) than did [n] syllables ([u] > [i] > [a]). In the talker analysis, there was also a significant talker sex by cutback interaction, $F(7,28) = 2,74$, $p = .0268$, reflecting a reduction of the talker sex effect as the syllables were progressively truncated. A separate analysis of the isolated murmurs only (cutpoints +1, +2, +3) showed neither the consonant main effect nor the consonant by vowel interaction to be reliable, due to large talker variability.

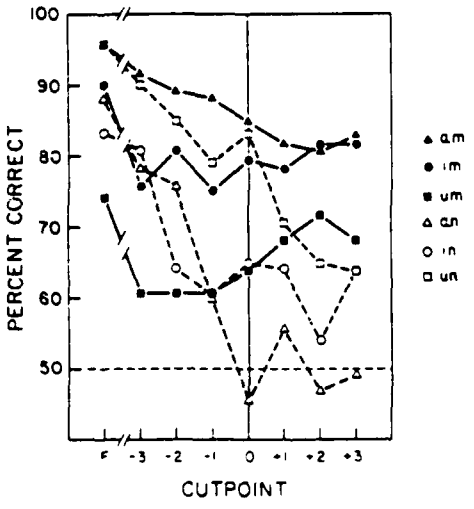


Figure 3. Identification scores for individual syllables in the Murmurs condition (Experiment 1).

Vowels

The average results of the Vowels condition are plotted in Figure 2 as the filled triangles. The score for full syllables is on the right-hand side here, and it is even lower than that for the identical stimuli in the Murmurs condition (80% correct). This poor accuracy was in part due to two subjects who, for unexplained reasons, performed at chance level with the full syllables, even though they did all right in the subsequent stimulus blocks of the Vowels condition. If their data are omitted, the score rises to 85% correct.

Reading the graph in Figure 2 from right to left, we see that elimination of all but 20 ms of the murmur (+2 cutpoint) left intelligibility unaffected, as it did also in the CV syllable experiment (Repp, 1986). Further truncation reduced identification performance gradually to 68% correct when the last 30 ms of the vowel were removed. Although performance seems to level off there, it presumably would have declined further, had the vowel been cut back more. The intelligibility score for truncated vowels (-3 cutpoint) is comparable to that for isolated murmurs, and also to that for truncated vowels and isolated murmurs of CV syllables (Repp, 1986). Although it is difficult to compare results across experiments, the prediction that the formant transitions of VC syllables would be relatively more informative than those of CV syllables is not supported.

The data for individual syllables are shown in Figure 4. Two syllables are clearly separated from the others here: Identification of [in] and [um] was poor to begin with and went to chance after elimination of the murmur. Not surprisingly, these are the syllables with minimal formant movements, due to the closeness of the places of articulation of vowel and consonant (cf. Repp, 1987). Truncated [a] vowels yielded the highest consonant identification scores, presumably because they have the most pronounced formant transitions. The pattern is quite different from that for CV syllables (Repp, 1986): Both [u(m)] and [u(n)] were much more poorly identified than their CV counterparts, whereas [i(m)] was identified better than [(m)i].

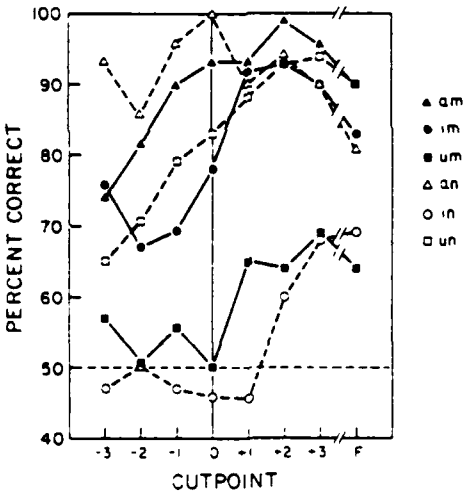


Figure 4. Identification scores for individual syllables in the Vowels condition (Experiment 1).

The statistical analysis revealed significant main effects of cutpoint, $F(7,77) = 8.35$, $p < .0001$; $F(7,28) = 8.70$, $p < .0001$, and of vowel, $F(2,22) = 98.93$, $p < .0001$; $F(2,8) = 20.97$, $p = .0007$, reflecting the higher performance for [a-] syllables, as well as an interaction between these two factors, $F(14,154) = 2.71$, $p = .0014$; $F(14,56) = 2.65$, $p = .0050$, due to the increase in the vowel effect as stimulus duration decreased. There was also a vowel by consonant interaction, $F(2,22) = 56.91$, $p < .0001$; $F(2,8) = 15.26$, $p = .0019$, reflecting the fact that [um] < [un] but [im] > [in], and a three-way interaction of these two factors with talker sex, $F(2,8) = 10.08$, $p = .0065$, since the vowel by consonant interaction was mainly due to the female talkers. With the murmurs of all stimuli cut back, it is difficult to attribute any remaining sex differences to the

removal of final releases in female tokens, so they must have a different origin (see Experiment 2).

Excerpts

The overall results for the Excerpts condition are shown as the open triangles in Figure 5. The left panel shows the effect of reducing the duration of excerpts from about 60 to 20 ms; the right panel shows the effect of varying the location of a 20 ms excerpt. As can be seen, performance for 60 ms excerpts was already poor but was reduced further when excerpt duration was shortened. Performance for 20 ms excerpts was only slightly above chance and apparently did not depend on location. These results contrast with those for CV syllable excerpts of the same duration, for which performance was not only generally higher but also showed an advantage for short excerpts straddling the murmur-vowel boundary. The local spectral change across the closure point does not seem to be perceptually important in VC syllables.

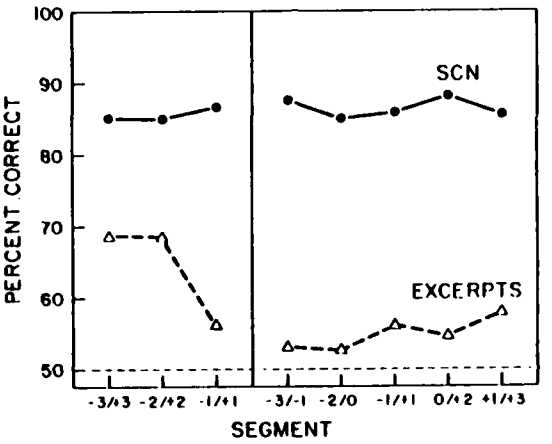


Figure 5. Percent correct identification for syllables in the SCN and Excerpts conditions (Experiment 1) as a function of SCN or excerpt duration (left panel) and location (right panel).

Results for individual syllables are shown in Figure 6. Scores for [un] and [in] were lower than those for the other syllables. Otherwise, the general conclusions hold for individual syllables. Only one syllable, [an], showed a tendency for a peak in the center of the right-hand graph. Note that the pattern for vowel excerpts (-3/-1) approximates that for truncated vowels (Figure 3), whereas that for murmur excerpts (+1/+3) resembles that for isolated murmurs (Figure 4), with the characteristic advantage for [m] murmurs.

The ANOVAs on the three conditions of varying duration showed the expected main effect of duration, $F(2,22) = 7.67$, $p = .0030$; $F(2,8) = 11.01$, $p = .0050$, as well as a vowel main effect, [a] > [i,u], $F(2,22) = 25.98$, $p < .0001$; $F(2,8) = 10.15$, $p = .0064$. The analysis of the five 20 ms excerpt conditions varying in location yielded main effects of consonant, $F(1,11) = 5.95$, $p = .0329$; $F(1,4) = 16.82$, $p = .0148$, and vowel, $F(2,22) = 6.48$, $p = .0061$; $F(2,8) = 7.00$, $p = .0175$, but no effect of location.

SCN

The overall results of the SCN condition are shown in Figure 5 as the filled circles. The conditions in the left panel vary in SCN duration, whereas those in the right panel vary in SCN location. Neither variable, however, seemed to have much effect on performance, which was similar to that for full, unaltered syllables. The results for individual syllables are shown in Figure 7. These functions, too, are rather flat,

and the differences among syllables are similar to those found among full syllables (see Figures 3 and 4). These results contrast with those for CV syllables, where perception of [mi] and [ni] was strongly affected by SCN.

ANOVAs on the duration data revealed no significant effect of that factor, only a vowel by consonant interaction, $F(2,22) = 45.56$, $p < .0001$; $F(2,8) = 6.63$, $p = .0200$, due to [am] > [an], [im] > [in], but [um] < [un]. Likewise, in the ANOVAs on the location data there was no main effect of location but the same vowel by consonant interaction, $F(2,22) = 28.69$, $p < .0001$; $F(2,8) = 11.38$, $p = .0046$, plus a three-way interaction of these two factors with talker sex, $F(2,8) = 5.90$, $p = .0266$; as usual, the differences were more pronounced for female talkers.

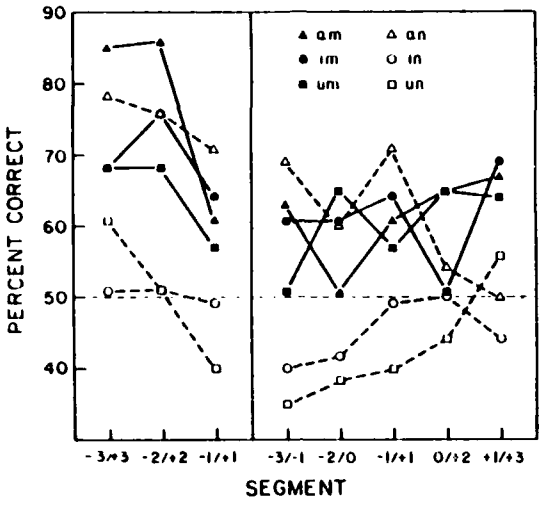


Figure 6. Identification scores for individual syllables in the Excerpts condition (Experiment 1).

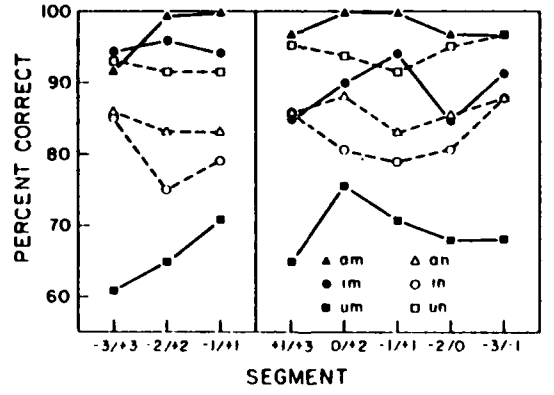


Figure 7. Identification scores for individual syllables in the SCN condition (Experiment 1).

Summary and Conclusions

Even though full VC syllables (with releases removed) were less intelligible than CV syllables, identification scores were about the same for their respective components (murmur and vowel) presented in isolation. If the stimulus components were equally informative, why were full VC syllables less well identified than full CV syllables? The reason may be that relational spectral information is less salient in VC stimuli, because of the relatively more gradual transition from the vowel to the

murmur. Listeners then have to rely mainly on integrating the information provided independently by vowel and murmur in VC syllables, whereas in CV syllables the relatively abrupt change in spectral structure across the release serves as an additional salient cue. Several observations are consistent with this suggestion. First, there was no perceptual advantage for brief excerpts that straddled the point of closure; thus, in contrast to CV syllables, having the point of maximal spectral change available did not aid identification.

Second, substituting SCN for as much as 60 ms of waveform surrounding the closure point left performance virtually unaffected, suggesting that listeners continued to integrate information provided by (truncated) vowel and murmur across the noise. In CV syllables, by contrast, identification of the [mi]-[ni] distinction was severely affected by interpolated SCN. Third, the [im]-[in] distinction was generally perceived more accurately than the [mi]-[ni] distinction, due to much better identification of the labial consonant in VC syllables. There was reason to believe that relational spectral information is especially crucial for [mi] identification, whereas it seems to be much less important for [im] identification. (Why the *isolated* signal components are more informative for [im] than for [mi], however, remains a mystery.) Finally, Repp (1986) proposed a simple mathematical model of cue integration that predicts exactly the obtained overall score of 85% correct for full syllables, given two independent cues that lead, by themselves, to 70 and 75% correct identification, respectively (cf. Figure 2). Thus, overall at least, VC identification performance seems more consistent with the hypothesis of independent cues than does CV identification performance.

The hypothesis that relational spectral information for place of articulation in nasals is less important in VC than in CV syllables will be pursued further in Experiment 4.

EXPERIMENT 2

One potential problem with Experiment 1 (and with Experiments 3 and 4, which used the same stimulus materials) is the inclusion of both unreleased and originally released tokens, with their releases removed. The latter tokens were produced almost entirely by female talkers. Female speech turned out to be less intelligible than male speech. This could have been a genuine effect of talker sex--an interesting ancillary finding. Alternatively, however, the removal of the final release may have made originally released tokens less intelligible than originally unreleased ones, either because there are articulatory and acoustic differences between released and unreleased tokens preceding the release, or because of some waveform-editing artifact (e.g., abrupt offset of the murmur). This would constitute a methodological flaw.

Experiment 2 was a control study conducted with newly recorded materials after Experiments 1, 3, and 4 had been completed. Its purpose was to determine (1) whether removal of final releases reduces intelligibility (as it most likely will), (2) whether the resulting release-less tokens are less intelligible than unreleased tokens produced by the same talkers, and (3) whether female tokens are less intelligible than male tokens.

Methods

Ten new talkers, five males and five females, were recorded producing the same VC syllables as in Experiment 1. Each talker was instructed to produce the syllables first with a final release (as demonstrated by the experimenter) and then without a release. All syllables were digitized, and their waveforms were inspected and carefully listened to, to make sure all tokens had been produced as intended. (The recordings of

several additional talkers were rejected because they seemed to contain ambiguous tokens; this was especially evident for one female.) The closure point was marked in all syllables, following the same procedures as in Experiment 1. The final release portion of the released (R+) tokens was located by eye and ear and removed to generate a set of release-less (R-) tokens, to be compared with the originally unreleased (UR) syllables.

Three randomized test sequences were recorded. The first contained the full syllables (10 talkers x 6 syllables x 3 versions = 180 stimuli), the second the vowels only (120 stimuli, since R+ and R- tokens were identical here), and the third the murmurs only (180 stimuli), with interstimulus intervals of 2.5 s. Ten subjects identified the stimuli as /m/ or /n/.

Results and Discussion

The results are displayed in Table 1. Percent correct scores are broken down by test (full syllables, vowels, murmurs), talker sex (male, female), release type (R+, R, UR), and syllables. The right-most column lists the scores for all syllables combined.

TABLE 1

Percent Correct Identification of Individual Syllables in Experiment 2 (R+ = Released, R- = Release-less, UR = Unreleased).

Sex	Type	[im]	[in]	[ɑm]	[ɑn]	[um]	[un]	Average
Full syllables								
Male	R+	100	100	100	98	100	100	100
	R-	98	98	100	98	90	100	97
	UR	98	78	100	96	96	92	93
Female	R+	90	96	100	98	96	100	97
	R-	94	92	100	90	60	88	87
	UR	94	92	100	100	56	84	88
Vowels								
Male	R	78	60	94	96	62	94	81
	UR	86	48	92	100	56	94	79
Female	R	66	50	92	98	46	74	71
	UR	70	50	70	96	38	76	67
Murmurs								
Male	R+	98	86	98	70	100	76	88
	R-	88	74	78	70	80	70	77
	UR	66	48	90	58	62	50	62
Female	R+	82	100	98	92	100	96	95
	R-	58	86	78	68	50	76	69
	UR	60	82	70	80	42	86	70
	UR	60	82	70	80	42	86	70

The average scores for full syllables show that removal of the releases from released tokens resulted in an intelligibility decrease of about 3% for male and 10% for female talkers. (Because of ceiling effects, the decrement was not tested for significance.) Importantly, however, the R- tokens were no less intelligible than the UR tokens; in fact, they received slightly higher scores in male productions (a

nonsignificant difference). In addition, male syllables were more intelligible than female syllables, though this difference was not quite significant across talkers, $F(1,9) = 32.01$, $p = .0003$; $F(1,9) = 4.45$, $p = .0679$. Inspection of scores for individual talkers revealed that three female talkers were much less intelligible than the remaining seven talkers. Thus, the conclusion is warranted that some, but not all, female-produced syllables were more ambiguous than male-produced syllables. Moreover, Table 1 shows that the sex difference for full syllables rests almost entirely on [u-] syllables.

Similar results were obtained for the isolated vowel portions. There was no significant difference between released and unreleased syllables; if anything, the former were slightly more intelligible. However, there was a significant talker sex effect in favor of male speech, $F(1,9) = 36.04$, $p = .0002$; $F(1,9) = 11.85$, $p = .0088$. This difference was more or less present for all six syllable types.

Removal of the releases from isolated released murmurs resulted in a substantial intelligibility decrement, $F(1,9) = 100.83$, $p < .0001$; $F(1,9) = 26.19$, $p = .0009$, especially in the female tokens, although the two-way interaction was not significant across talkers. As to the comparison between R- and UR tokens, the former were the more intelligible in male speech, though the corresponding main effect and interaction were not significant across talkers. There was no significant talker sex effect for isolated murmurs.

The average scores for R- and UR tokens combined across all talkers were very similar to those obtained with analogous stimuli in Experiment 1: 91% correct here versus 88% there for full syllables; 74% versus 75% for vowels; 70% versus 70% for murmurs.

In summary, these results vindicate the procedures of Experiment 1. Although removal of the release from released final nasal consonants does impair their intelligibility, the resulting release-less tokens are not less intelligible than tokens produced without any release. Rather, there seems to be a genuine difference in the intelligibility of male and (some) female talkers, which derives from the vowel rather than the murmur portion of the stimuli, as already suggested in Experiment 1. The reason for this difference is unknown; one possibility is that anticipatory vowel nasalization is especially strong in some female talkers, and that this makes the formant transitions less salient. We may now proceed to Experiments 3 and 4, which used the materials of Experiment 1.

EXPERIMENT 3

The purpose of Experiment 3 was to examine one particular explanation of the perceptual contribution of the nasal murmur to place of articulation perception (cf. Repp, 1987: Experiment 5). In addition to providing direct spectral cues to place of articulation, the murmur serves as an important, and possibly essential, manner cue. Failure to perceive the correct nasal manner may interfere with place of articulation perception, since the precise acoustic structure of the vocalic formant transitions may deviate from that typically associated with oral stop consonants. Initial nasal consonants are often perceived as nonnasal stops when the murmur is removed (Kurowski & Blumstein, 1984; Repp, 1987), but for a stimulus of ambiguous manner, place of articulation identification is no better when manner is identified correctly than when it is not (Repp, 1987). This suggests that accuracy of place perception does not depend on oral/nasal manner perception in CV syllables. Experiment 3 examined the same issue in VC syllables. Because of the stronger nasalization of the vowel in this context, it was expected that removal of the murmur would interfere less with perception of nasal manner than it does in CV syllables.

Methods

Experiment 3 was run in two different versions, which will be referred to as 3a and 3b. The Vowels tape of Experiment 1 was reused in Experiment 3a. That is, the subjects heard 8 sequences of 36 stimuli each, and the syllables were truncated progressively from the end. The only difference was in the instructions: Whereas previously the subjects had to make a forced choice between /m/ and /n/, they were now encouraged to write down /m,n,b,d/ or any other consonant they heard, or a dash if they did not hear any consonant. A new group of 12 paid student volunteers was recruited.

Experiment 3b was motivated by the suspicion that the subjects in Experiment 3a may have been biased against stop consonant responses by hearing full nasal syllables first. The instructions of Experiment 3b emphasized even more that responses should reflect what was heard, not inferences about deleted consonants. Only stimuli truncated at points 0, -1, -2, and -3 (i.e., vowels without murmurs) were used in a completely randomized sequence. A new group of eight subjects served as listeners.

Results and Discussion

The overall results are shown in Figure 8 in terms of three response measures: the percentage of consonant responses, p(C); the percentage of correct place of articulation identifications given that a consonant was heard, $p_c(P|C)$; and the percentage of nasal consonant responses given that a consonant was heard, p(N|C). Consonant responses in Experiment 3a dropped from 100% to about 80% as the syllables were cut back. In Experiment 3b, fewer consonant responses were given, and truncation of the vowel seemed to have a stronger effect than in Experiment 3a. The place identification scores were rather similar in the two experiments, falling from an initial 90% correct in full syllables (Experiment 3a) to about 75% correct at cutpoint -3. (Considering that, in a two-alternative forced-choice task, random guesses on the remaining trials would increase the score to about 88% correct, the present subjects' performance was much better than that of the subjects in Experiment 1, for reasons that are not clear.) Nasal consonant responses dropped slightly with truncation in Experiment 3a, to about 88%, but more sharply in Experiment 3b, to 71%. This difference was probably a consequence of the changes in instruction and stimulus sequence.

Of the differences among individual syllables, the following are worth mentioning: (1) Consonant responses were less frequent for [in] and [um] than for the other syllables. A similar difference was observed for CV syllables by Repp (1987). It reflects the shallow formant transitions in these syllables, which have similar articulatory configurations for vowel and consonant. (2) Perception of nasal manner was poorest in [im] and [in], just as in CV syllables (Repp, 1987), probably because high vowels are generally less nasalized. Nasal consonant responses dropped by a substantial amount for these two syllables in Experiment 3b. Even so, each VC syllable in Experiment 3b received more nasal responses than the corresponding CV syllable in the analogous experiment of Repp (1987: Experiment 5), which almost certainly reflects the greater nasalization of vowels preceding rather than following a nasal consonant.

It is evident from these results that removal of the nasal murmur interfered with the perception of nasality, but not very much. Therefore, only a small part of the improvement in place of articulation perception consequent upon the reinstatement of a deleted murmur could be due to the full restoration of nasal manner cues. An additional response measure that bears on this issue is the correct place of

articulation score contingent on perceived nasal or nonnasal manner. If correct place perception depended on correct manner perception, then place identification should be more accurate in nasal than in oral consonant responses to the same stimuli. These percentages, computed from the average scores for each syllable at cutbacks 0, -1, -2, and -3 and subsequently averaged over syllables and truncation conditions, were 76.2% (nasal) and 81.8% (oral) in Experiment 3a and 74.5% (nasal) and 76.0% (oral) in Experiment 3b. The latter values are more reliable because of the larger proportion of nonnasal responses in Experiment 3b. Neither, however, give any indication that perception of nasal manner aided place of articulation identification. A similar conclusion was reached for CV syllables (Repp, 1987).

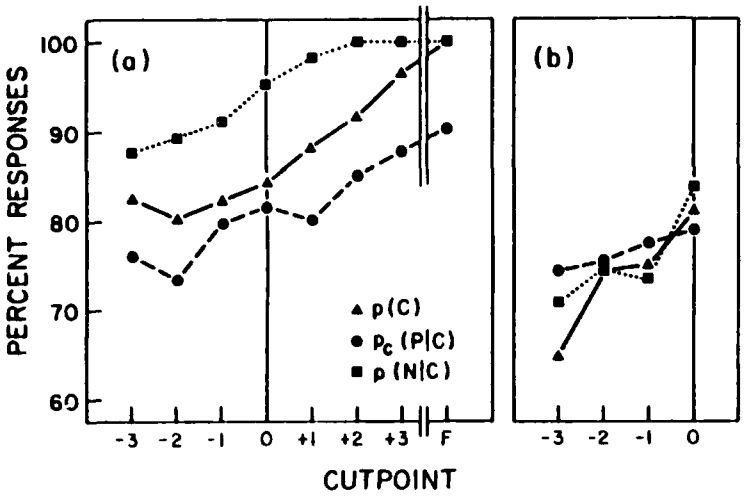


Figure 8. Percentage of consonant responses (triangles), of correct place of articulation identifications given that a consonant was heard (circles), and of nasal responses given that a consonant was heard (squares). Panels a and b show results of Experiments 3a and 3b, respectively.

In both parts of the experiment, listeners gave more labial responses when identifying stimuli as nonnasal rather than nasal, but this difference (5% and 10%, respectively, in Experiments 3a and 3b) was much smaller than that obtained for CV syllables (Repp, 1987). For CV syllables, this criterion shift was attributed to the absence of stop consonant release bursts in prevocalic nasals, but this consideration does not apply to postvocalic nasals. Part of the bias, therefore, seems to have a different origin, perhaps deriving from differences between oral and nasal stop consonants in formant trajectories and/or vowel amplitude envelope.

In summary, this experiment suggests that the contribution of the murmur to place of articulation perception derives directly from spectral information in the murmur (i.e, its formants above 1 kHz), not indirectly from nasal manner cues. This outcome is consistent with the hypothesis that, in VC syllables, vowel and murmur constitute independent sources of place of articulation information.

EXPERIMENT 4

Experiments 4 had two parts that were run in a single experimental session that included also Experiment 3a. Their purpose was to obtain further evidence on the relative (un)importance of relational spectral information in the perception of postvocalic nasals. Each part resembled an earlier experiment with CV syllables (Repp, 1986, 1987).

Experiment 4a

This experiment corresponds to Experiment 2 in Repp (1987). In that study, a decrement in CV identification performance was found when the murmur and vowel components were presented successively to opposite ears. This decrement was attributed to subjects' increased difficulty in utilizing relational spectral information across different channels. Since it has been hypothesized that such relational information is less important for the perception of VC syllables, the prediction is that channel separation will prove relatively less harmful to VC syllable identification. That is, even when vowel and murmur occur in different ears, listeners should be able to integrate these two sources of information centrally, and this may be all that is needed to identify final nasal consonants.

Methods

This short test consisted of 108 stimuli resulting from the randomization of three conditions with 36 syllables each. In one condition (V+M), vowel and murmur immediately followed each other on the same channel. In the second condition (V/M), the vowel occurred on one channel and the murmur on the other. In the third condition, the isolated vowel was presented without the murmur. All vowel portions occurred on the same channel, which was presented to the left ear for half of the subjects and to the right ear for the other half. The subjects made a forced choice between /m/ and /n/. Since no ear differences were apparent, the data of all subjects were combined.

Although it was not strictly necessary, a procedure used in the corresponding CV syllable experiment to avoid ceiling effects was followed here also: Both vowel and murmur were truncated by 30 ms (i.e., at cutpoints -3 and +3, respectively). Actually, this manipulation worked against the hypothesis under test: The truncation introduced a more abrupt spectral change between vowel and murmur than occurs normally in VC syllables.

Results and Discussion

The results are shown in Table 2. The overall score for the V+M syllables was 81% correct, and that for isolated vowels was 67% correct, in agreement with Experiment 1. Performance in the split (V/M) condition was 77% correct, only slightly lower than in the V+M condition. In fact, this difference was not significant, whereas that between the V/M and V conditions was, $F(1,11) = 17.03$, $p = .0017$; $F(1,5) = 8.33$, $p = .0344$. The pattern of differences among individual syllables is consistent with earlier results (see Figure 4), except for the poor identification of [im] in isolated vowels.

TABLE 2

Percent Correct Identification of Individual Syllables in Experiment 4a (V = Vowel, M = Murmur, / = Split Channels).

Condition	[im]	[in]	[am]	[an]	[um]	[un]	Average
V+M	86	75	96	86	53	89	81
V/M	83	64	90	86	65	76	77
V	56	47	81	94	57	64	67

These results are consistent with the prediction that channel separation of vowel and murmur has little effect on VC intelligibility. For CV syllables, by contrast,

performance was significantly (8%) lower in the M/V than in the M+V condition (Repp, 1987). The results thus support the hypothesis that the identification of final nasal consonants rests mainly on the integration of independent cues, with no significant contribution of relational spectral information.

Experiment 4b

This experiment corresponds to Experiment 3 in Repp (1987) on CV syllables. It was intended to address the same hypothesis as the preceding experiment, viz., that spectral change information is relatively unimportant in VC syllables, but used a different technique to separate the murmur from the vowel---intervals of silence. The prediction was that temporal separation of the two signal components, within limits, should have little effect on intelligibility.

Methods

The vowel and murmur components were the same as in the preceding experiment; that is, they had been truncated by 30 ms to increase the error rate. Silent intervals of five durations were inserted between the two components: 0, 30, 60, 120, and 240 ms. Thus there were 5 x 36 = 180 stimuli, which were all randomized together. The subjects made a forced choice between /m/ and /n/ for each stimulus.

Results and Discussion

The results are shown in Table 3. It is evident that the effect of separating vowel and murmur by varying amounts of silence was minimal, as predicted. The effect of silence duration was nonsignificant in both ANOVAs. Although the effect of a similar manipulation on CV syllables (Repp, 1987) was smaller than expected, it was at least twice as large as the present effect and significant across subjects. Thus the results lend further support to the hypothesis that relational spectral information is less important in VC than in CV syllables.

TABLE 3

Percent Correct Identification of Individual Syllables in Experiment 4b.

Silence	[im]	[in]	[ɑm]	[ɑn]	[um]	[un]	Average
0 ms	88	69	99	86	64	94	83
30 ms	85	81	94	92	60	92	84
60 ms	85	85	94	90	56	93	84
120 ms	85	75	90	86	65	92	82
240 ms	89	75	89	90	54	88	81

SUMMARY AND CONCLUSIONS

The present series of experiments provides a variety of evidence in support of the hypothesis that relational spectral information is less important for place of articulation identification of nasal consonants in VC than in CV syllables. This evidence includes (1) lower identification scores for full VC than CV syllables despite comparable intelligibility of the isolated components (Experiments 1 and 2), (2) absence of a peak in identification performance for brief excerpts straddling the closure point, and absence of a dip in performance when the same signal portions were replaced with SCN (Experiment 1), (3) absence of a significant intelligibility decrement in split-channel presentation of vowel and murmur (Experiment 4a), and

(4) absence of a significant effect of temporal separation (of up to 240 ms) of vowel and murmur (Experiment 4b). Although these results are mostly negative, they contrast with the stronger effects of similar manipulations in CV syllables. Since the spectral relationship between vowel and murmur seems to be unimportant in the perception of postvocalic nasals, and since the murmur does not appear to make its contribution to place of articulation perception via its function as a manner cue (Experiment 2), the vowel and murmur essentially provide two independent sources of spectral information that are presumably integrated by the listener at a cognitive level (see Massaro & Oden, 1980).

Why are spectral relationships less important in VC than in CV syllables? Two factors may play a role. First, as already mentioned in the introduction, spectral change is simply less pronounced and occurs at a slower rate in VC syllables, because of greater anticipatory nasalization, possible articulatory motion beyond the point of closure, and perhaps also slower closing than opening gestures. Second, there may be an inherent asymmetry in auditory sensitivity to direction of a change in level. The spectral change across the release in CV syllables is not only more abrupt but entails increases in the levels of most frequencies in the spectrum. The change in VC syllables across the point of closure, on the other hand, consists of a level reduction across most of the spectrum. Psychoacoustic research has shown that intensity decrements in pure tones are poorly detected by human infants and monkeys, even though adult humans detect them as well as intensity increments (Sinnott & Aslin, 1985; Sinnott, Petersen, & Hopp, 1985).

One might speculate that critical phonetic learning occurs in humans before the auditory capacities necessary for intensity decrement detection are acquired. Moreover, in a relatively continuous sound, auditory short-term adaptation may reduce a listener's sensitivity to relatively smooth intensity decrements. Paradoxically, then, adaptation may not be responsible for the relative importance of relational spectral information in CV syllables (as hypothesized by Kurowski and Blumstein, 1984, but called into question by Repp, 1987) but rather for its lack of importance in VC syllables.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. We are grateful to Diana Matson for extensive assistance with Experiment 2, to Hwei-Bing Lin for assistance with Experiment 3b, and to all colleagues at Haskins Laboratories who served as talkers.

REFERENCES

Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. (1971). Perception of coarticulated nasality. *Journal of the Acoustical Society of America, 49,* 538-540.

Broad, D. J., & Fertig, R. H. (1970). Formant-frequency trajectories in selected CVC-syllable nuclei. *Journal of the Acoustical Society of America, 47,* 1572-1582.

Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, 76,* 383-390.

Kurowski, K., & Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America, 81,* 1917-1927.

Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language , 32,* 274-284.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America, 67,* 996-1013.

Miller, J. L. (1977). Nonindependence of feature processing in initial consonants. *Journal of Speech and Hearing Research, 20,* 519-528.

Ohde, R. N., & Sharf, D. J. (1977). Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. *Journal of Speech and Hearing Research, 20,* 543-554.

Ohde, R. N., & Sharf, D. J. (1981). Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study. *Journal of the Acoustical Society of America, 69,* 297-300.

Ostreicher, H. J., & Sharf, D. J. (1976). Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics , 4,* 285-301.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32,* 693-703.

Pols, L. C. W., & Schouten, M. E. H. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America, 64,* 1333-1337.

Pols, L. C. W., & Schouten, M. E. H. (1981). Identification of deleted plosives: The effect of adding noise or applying a time window (A reply to Ohde and Sharf). *Journal of the Acoustical Society of America, 69,* 301-303.

Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. *Journal of the Acoustical Society of America,* 73, 1346-1353.

Repp, B. H. (1986). Perception of the [m]-[n] distinction in CV syllables. *Journal of the Acoustical Society of America, 79,* 1987-1999.

Repp, B. H. (1987). On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast. *Journal of the Acoustical Society of America, 82,* 1525-1538.

Schouten, M. E. H., & Pols, L. C. W. (1979). CV- and VC-transitions: a spectral study of coarticulation--Part II. *Journal of Phonetics, 7,* 205-224.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735-1736.

Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78,* 1986-1992.

Sinnott, J. M., Petersen, M. R., & Hopp, S. L. (1985). Frequency and intensity discrimination in humans and monkeys. *Journal of the Acoustical Society of America, 78,* 1977-1985.

FOOTNOTES

**Journal of the Acoustical Society of America,* in press.

†Also Department of Linguistics, University of Connecticut.

Orchestrating Acoustic Cues to Linguistic Effect*

Leigh Lisker†

A most convincing way to demonstrate that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with respect to that property, but in nature such pairs are most unlikely. The English words rapid and rabid are a minimal pair at the level of the segmental phoneme, and are near minimal at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. Three properties lend themselves to simple waveform editing: the duration of the stressed vowel, the duration of the closure, and the glottal buzz versus silence of the closure signal. Listener responses to stimuli having natural values of these properties show that, with a single exception, there was no decisive effect on word identification produced by a shift in the value of any one property. Adding glottal buzz to the /p/ closure led listeners to report the word "rabid." To transform original "rabid" to "rapid," at least two properties had to be changed to achieve any significant effect, one of these necessarily the replacement of closure buzz by silence.

Phonetic research nowadays considers the processes involved in speech communication from a wide variety of perspectives, but a central concern remains that of identifying and characterizing those features of the speech processes that serve a message-differentiating function. The phonetic analysis of a speech signal into a temporal sequence of sounds, as well as the decomposition of those sounds into features, provide a framework within which to specify the distinctive properties that determine a particular interpretation of the signal. A coherent account of a given speech event, considered as representative of a set of linguistically identical events, states the mutual interrelations among physiological, anatomical, and acoustic patterns, and how they relate to the listener responses they elicit. By far the greatest attention has been given to finding the acoustic cues to the linguistic message conveyed by a speech signal. The search has involved the analysis of speech signals, the selection of promising cue candidates, and the experimental assessment of their cue value by the methods of speech synthesis and tests of perception. Such evaluation of a feature's cue value typically has involved the use of acoustic patterns designed to maximize the likelihood that the feature of interest will affect listeners' response behavior. The number of acoustic pattern features that have been determined to have cue value is not known with certainty, and presumably with continued research along established lines that number will only increase. Clearly it is easier to show that a feature has cue value than to justify a claim to the contrary (the famous unprovability of the null hypothesis).

Most of the acoustic cues so far uncovered are referred to as segmental cues, or even cues to particular phonetic features of segments. The experimental data supporting their identification are derived by means of some variant of the linguist's "minimal pair" test. A most convincing way to demonstrate that an acoustic property is a cue for the listener would be to find speech events that constitute minimal pairs with

respect to that property, but in nature such pairs are unlikely. The English words *rapid* and *rabid* make a minimal pair at the level of the segmental phoneme, and almost minimal at the level of the phonetic feature, but as many as sixteen acoustic properties are candidate cues to the lexical distinction. It is not certain, however, that any one of them is an independent cue, that is, one that can signal a lexical distinction independently of its acoustic context. Furthermore, even if a given property can be shown to have such power to affect listener responses, it need not be true that this property functions independently in natural speech.

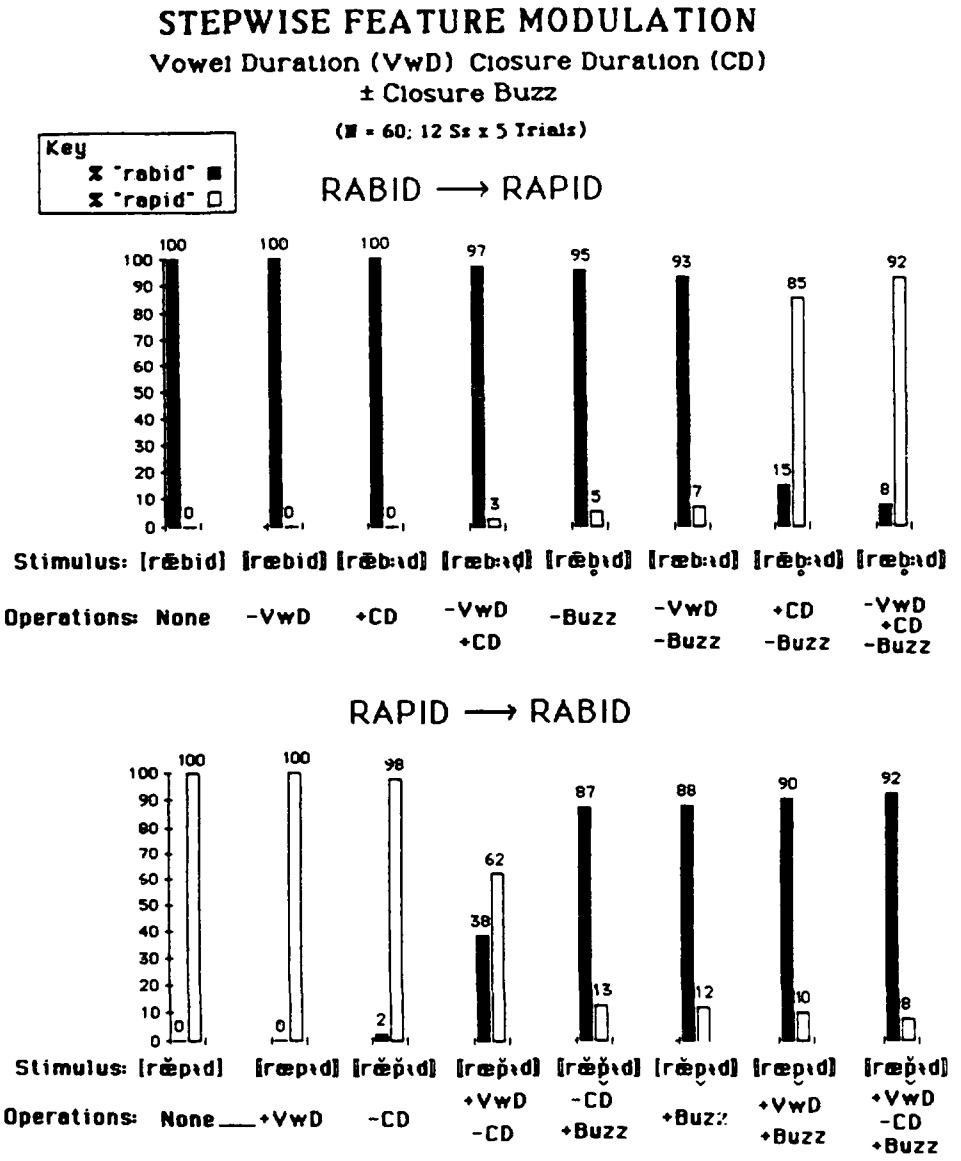
In the following I want to report some listener responses to sets of stimuli derived by waveform editing of some naturally produced tokens of *rapid* and *rabid*. Three properties served as experimental variables: the closure duration interval, the glottal buzz/silence difference during closure, and the duration of the steadystate [æ] vowel. Unlike many tests of this kind, in which the values assigned a variable are altered in steps of a size designed to establish one or more category boundaries, in the test here reported each variable was given just two (for one variable three) values, these being chosen on the basis of naturalness. This does not mean that the stimuli thus derived can be called natural, only that the values assigned the experimental variables were found in natural utterances recorded by the talker who served as the signal source.

From a set of recordings of the expressions *I think it's rapid* and *I think it's rabid* produced by a male speaker of a central East Coast variety of American English a typical token of each sentence was selected, digitized, and stored on computer. A waveform editing program was applied to produce a total of sixteen stimuli having acoustic differences restricted to the intervals corresponding to the final words of the utterances. The durations of the acoustic intervals corresponding to the labial closures were given two values, 60 and 115 ms, these being values typical of the talker's productions of the /p/ and /b/ closures in the particular context. The closure intervals were either acoustically blank or filled with buzz derived from the original /b/ closure. The pre-closure intervals, from the cessation of the noise interval marking the /s/ preceding the target word to the beginning of the labial closure, were set to the following values: for derivatives of *rabid* the mean /b/ value of 270 ms and a shorter duration of 230 ms; for *rapid* derivatives the mean /p/ value of 190 ms and an increased duration of 230 ms. The common value of 230 ms was chosen because it fell within the range of natural values for both words in the sentence context used. (Shortening the pre-closure span of *rabid* to 190 ms, the mean /p/ duration, effected a noticeable shift in vowel quality.) A test order in which each of the sixteen stimuli was presented five times, that is, a random order of eighty items, was presented to twelve native American English speakers, all linguistically and phonetically naive. Each test item was composed of an acoustically invariant carrier *I think it's* followed by the target word to be identified.

The figure illustrates the perceptual effects of the several shifts in the values of the three acoustic properties. The upper panel represents changes in the perception of the original *rabid* token, while the lower panel shows the effect of editing the original *rapid*. For each of the variables a change to a value not normally associated with the original stimulus type has, with one exception, no great effect on labeling behavior. Only when glottal buzz replaces the silence of the /p/ closure is there a decided shift to "rabid" judgments.

It does not follow, of course, that the three variables are otherwise of negligible importance for the perception of the two words. Thus a combination of devoicing and lengthening of the /b/ closure elicited an overwhelmingly "rapid" response, a result in conformity with earlier findings. But a shortening of the /p/ closure together with a lengthening of the preceding vocalic interval still yielded mostly "rapid" judgments. Original "rapid" was heard largely as "rabid," while "rabid" went to

"rapid," when all three variable features were assigned values appropriate to the competing form.



The results summarized above indicate that changing the value of an acoustic feature to which cue value has been attributed does not always produce a significant effect on linguistic labeling behavior; its effect is quite context-dependent. Indeed it may well be, in the case of certain properties, that the context in which they can be decisive can only (?) be contrived in the laboratory. The status of an acoustic property of speech is therefore very different from that of a phonetic feature, which we generally suppose to possess the power, for at least some natural phonetic systems, to

mark differentially some words from others, and to do this independently of other phonetic features.

ACKNOWLEDGMENT

This work was supported by NICHD Grant HD-01994 to Haskins Laboratories.

FOOTNOTES

*This text is a slightly modified version of a paper presented at the 11th International Congress of Phonetic Sciences, held 1-7 August 1987 in Tallinn, Estonia, U.S.S.R., and published in the *Proceedings XIth ICPhS, Vol. 6, 66-67,* Tallinn 1987.

†Also University of Pennsylvania

Book Review*
(Review of Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press)

Arthur S. Abramson†

It is not easy to know what to make of this book for people who read our journal and attend our meetings. For such people, surely, the science of phonetics rests upon a foundation of physiological and acoustic research, as well as psychological testing of hypotheses on the information-bearing elements of the acoustic signal and their underlying articulatory mechanisms. Readers with this outlook may find themselves feeling uncomfortable over the author's eclectic use of mainly impressionistic phonetic statements from a wide variety of sources of, seemingly, varying levels of reliability. This is generally so even though Maddieson, who is certainly not unsophisticated in these matters, does occasionally draw upon instrumental or psychological research.

The book is not one to be read from cover to cover. Rather, it is a reference book based on the UCLA Phonological Segment Inventory Database (UPSID), which resembles in some respects the nearby Stanford Phonology Archives (SPA). UPSID contains 317 languages, one from each major subgoup of each language family. This genetic sampling is meant to be typologically representative.

There are 10 chapters: 1. The size and structure of phonological inventories, 2. Stops and affricates, 3. Fricatives, 4. Nasals, 5. Liquids, 6. Vocoid approximants, 7. Glottalic and laryngealized consonants, 8. Vowels, 9. Insights on vowel spacing (contributed by Sandra Ferrari Disner), and 10. The design of the UCLA Phonological Segment Inventory Database (UPSID). There are two appendices that make up more than half of the book: A. Language lists and bibliography of data sources, and B. Phoneme charts and segment index for UPSID languages.

The book is really meant for linguists who need a statistically reliable base for the discovery of generalizations about phonological inventories that will be useful "in the formulation of phonological theories, in evaluating competing historical reconstructions, in constructing models of language change and language acquisition..." and can stimulate "important linguistically-oriented phonetic research" (p.1). Indeed, even without computer-access to UPSID itself, the linguist can make use of the well-planned organization of the book for such goals. Of course, the speech scientist who is not also a card-carrying linguist may well be interested in at least the last use mentioned. For the sake of this review, I have done a simple little test of the book by raising two questions that a speech scientist might ask. One is about how many languages, if any, exploit a given possible mechanism. The other concerns the accuracy of the phonetic statements taken from the literature.

Haskins Laboratories

In connection with questions of excitation-switching, one might want to know whether systematic use is made in any language of the possibility of moving from the local turbulence of a constriction in the supra-glottal vocal tract to continued noise-excitation of the vocalic formants of the relatively unimpeded tract upon release of the constriction, for some time before the onset of glottal pulsing as the next source. That is, just as there are aspirated stop consonants, are there also aspirated fricatives? (My hypothetical questioner finds the posited succession of events physiologically and acoustically plausible.) After a careful reading of Chapter 10 to learn the rules, one inspects the Segment Index (pp. 205-262) and finds, under Fricatives, an entry called Voiceless aspirated dental/alveolar sibilant fricative /"s^h"/ found in three languages: Burmese, Karen, and Mazahua. (Maddieson uses quotation marks around symbols to indicate imprecision in his sources as to exact place of ariculation.) One then consults the Alphabetic list of languages with key to sources to find the phoneme charts, where the consonant in question can be viewed in a paradigmatic array of all the phonemes arranged as intersections of largely traditional phonetic features. Thus, for example, in the Sgaw dialect of Karen described by R. B. Jones in 1961, Chart 516 shows aspirated /s^h/ is in contrast with plain /s/ and a rare voiced /z/, along with other fricatives. With only three languages in the data base showing this consonant type, we are further intrigued to find that two of them, Karen and Burmese, are genetically related in the Sino-Tibetan family, although they are in two branches of it, Karenic and Lolo-Burmese, respectively. (What we are not told is that there is extensive co-territoriality of these two languages in Burma.) The third language, Mazahua, is a member of the Oto-Manguean branch of "Northern Amerindian." The latter information has to be found by scanning the Genetic listing of languages and outline classification (pp. 174-177); there is no cross-referencing between the lists.

As a way of testing for accuracy, I studied Chart 400 on Standard Thai, a language I have worked on for a long time both impressionistically and instrumentally. I was dumbfounded to learn that Thai has a voiceless dental sibilant affricate /t̪s̪/ and a voiceless aspirated dental sibilant affricate /t̪s̪^{h}/. Turning to the alphabetic list, I find that the author's sources are studies by Noss in 1954 and 1964 and Abramson in 1962! Both Richard Noss and I follow Mary Haas in describing these plosives as voiceless unaspirated and aspirated palatal or, perhaps, palato-alveolar affricates. Both of us as Thai speakers of a certain fluency and, I believe, accuracy, would certainly agree that anything resembling a dental affricate would be a mispronunciation. Of course, Maddieson, a very reputable phonetician in his own right, could show Noss and me to be wrong, but he simply cites us without comment. I hasten to add, however, that casual inspection of the charts of other familiar languages does not reveal anything so egregious, although I am doubtful about some of the descriptive labels here and there. Cambridge University Press deserves no praise for its use of inelegant double-spaced pale typescript for this book. The available technology makes possible much clearer and darker single-spaced camera-ready copy.

Maddieson has produced a book that every speech laboratory will want to have as a handy reference. Insofar as one is willing to make allowances for the varying reliability of the many sources consulted to form the data base, the investigator can indeed test a wide range of hypotheses on phonemic patterning across a representative sampling of languages. The chapters in the first half of the book are interesting, well-written expositions of the topics and the many difficult obstacles encountered in a task of this size.

FOOTNOTES

*Journal of the Acoustical Society of America, 82, 720-721.

†Also University of Connecticut.

Appendix

SR #	Report Date	DTIC #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907

SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	**	ED 282-278
SR-89/90	January-June 1987	**	ED 285-228

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

END

DATE

FILMED

5-88

DTIC